THE UNIVERSITY OF TEXAS AT SAN ANTONIO, COLLEGE OF BUSINESS

Working Paper SERIES

Date February 12, 2016

WP # 0001MGST-253-2016

Performance Accuracy of Linear classifiers for Two-level Multivariate Observations in High-dimensional Framework

Anuradha Roy _
Department of Management Science and Statistics
The University of Texas at San Antonio
E-mail: Anuradha.Roy@utsa.edu

Tatjana Pavlenko
Departament of Mathematics
KTH Royal Institute of Technology
E-mail: pavlenko@math.kth.se

UTSA COLLEGE OF BUSINESS KNOWLEDGE FOR A NEW WORLD

ONE UTSA CIRCLE
SAN ANTONIO, TEXAS 78249-0631
210 458-4317 | BUSINESS.UTSA.EDU

Performance Accuracy of Linear classifiers for Two-level Multivariate Observations in High-dimensional Framework

Anuradha Roy *
Department of Management Science and Statistics
The University of Texas at San Antonio
One UTSA Circle
San Antonio, TX 78249, USA
Tel: +1-210-458-6343, Fax: +1-210-458-6350
E-mail: Anuradha.Roy@utsa.edu

Tatjana Pavlenko
Departament of Mathematics
KTH Royal Institute of Technology
SE-100 44 Stockholm, Sweden
Tel: +46-8-790-8466, Fax: +46-8-723-1788
E-mail: pavlenko@math.kth.se

Abstract

Asymptotic approximation for the expected probability of misclassification is derived for the linear classifier based on two-level multivariate data using block-compound symmetric (BCS) covariance structure. Advantages of this structure for modeling two-level multivariate data are shown in growing dimensions asymptotics which allows the number of replicates, u to grow faster than the total number of samples, n and only constraints the number of feature variables, $m < n$. Relevance and benefits of the designed classifier are demonstrated for a number of high-dimensional scenarios using both asymptotic results and simulations. Comparison of our findings with other existing results is considered.

Keywords Blocked compound symmetry; Two-level multivariate observations; Growing dimensions; Probability of misclassification.

Mathematics Subject Classification 62H30; 62G20.

1 Introduction

Performance accuracy of the classification procedures designed for high-dimensional data is extensively studied in the statistical literature. As the sample based classifiers usually have complex distributional expressions, various types of asymptotic approximations of the expected probability of misclassification (EPMC) are derived. Initially, the consideration was given to large sample theory, i.e., assuming that the number of feature variables, m is fixed and the total sample size, n goes to infinity. Okamoto (1963) in his renowned article first

*Correspondence to: Anuradha Roy, Department of Management Science and Statistics, The University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, USA

time derived the asymptotic expansion of the EPMC for linear classifier up to the second order term. Siotani (1982) also considered the large sample approximations and asymptotic behavior of a number of statistics related to classification accuracy. Later, Fujikoshi and Seo (1998) derived an asymptotic expansion of the same EPMC of the linear classifier in the high-dimensional framework, i.e., assuming that m can grow at the same rate as n. Very recently, in the same asymptotic settings Kubokawa et al. (2013) derived an expansion of EPMC for the ridge-type linear classifier up to the second order term and investigate the asymptotic effect of the ridge parameter on the classification accuracy. For the review of asymptotic expansions of the EPMC both in large-sample as well as in high-dimensional asymptotics, see Fujikoshi et al. (2010).

The above-mentioned asymptotic expansions of EPMC have been established for traditional multivariate data, or just for 'one-level multivariate data' in a high-dimensional setting. However, modern experimental techniques allow to collect and store *multi-level* (Leiva and Roy, 2011) high-dimensional data in almost all fields from agriculture to medical research, where the observations are collected on more than one feature variable (m) at different locations (u) repeatedly over time (v) and at different depths (d) etc. etc. These multi-level observations may have variances that differ across locations, time and depths, and efficient techniques for talking into account these variations is of great importance.

In recent years, there has been a flurry of activity over development of classification rules for multi-level multivariate data in high-dimensional setting, hitherto to the best of our knowledge no one has investigated their asymptotic performance properties. In a series of papers by Roy and Leiva (2007) and Leiva and Roy (2011, 2012) a number of linear and quadratic classifiers were developed for two- and three-level multivariate data in combination with structured mean and structured variance-covariance matrix. The main advantage of exploiting the structure underlying the data is shown in reducing the number of unknown parameters, which in turn leads to a significant gain in the classification accuracy. This advantage is especially important for small sample high-dimensional problems where the number of practically available observations is limited.

To investigate various effects of the covariance structure on the classification theoretically, one needs to derive asymptotic results on the *probability of misclassification* (PMC) in high-dimensions. In this article we obtain the asymptotic expansion of the EPMC up to the second order term for the linear classifier designed for two-level multivariate data in high-dimensional situation, which is an extension of the method proposed by Kubokawa et al. (2013) for one-level multivariate data.

To represent the two-level (m dimensional vector over u locations or time points) multivariate observations we use blocked compound symmetric (BCS) covariance structure (defined in Section 2.1) in high-dimensional setting and evaluate the performance of the corresponding classifier using the second order asymptotic approximation of EPMC. The designed classifier does not need the constraint $n > mu$, i.e., the total number of observations n larger than the total number of variables mu, it just needs $n > m$. In our high-dimensional asymptotics we assume that $m/n \to c$, where $0 < c < 1$, whereas u can go to infinity at the same rate as n.

It is interesting to note that the BCS covariance structure for two-level feature variables, which is a multivariate generalization of compound symmetry structure for multiple variables, has been initially introduced by Rao (1945, 1953) while classifying genetically different groups, and then it did not attract much attention in the literature for nearly half a century. Then Leiva (2007) developed classification rules for BCS covariance structure along with structured mean vectors in 2007. Lately, this covariance structure starts to gain a lot of attention in the literature, especially in the area of high-dimensional estimation and hypothesis testing (see Roy and Leiva, 2011; Coelho and Roy, 2013). An important advantage of using BCS structure is that the number of unknown parameters is only $m(m+1)$, which does not even depend on the number of repeated measures u, whereas the number of unknown parameters in the unstructured covariance matrix is $mu(mu+1)/2$, which can increase very rapidly with the increase of any one of the levels m or u. Hence, using BCS covariance structure, one can allow the number of repeated measurements u to grow unrestrictedly and thereby providing more information, while the number of unknown parameters remains the same.

This paper proceeds as follows. In Section 2, the BCS covariance structure is introduced, and estimation of its matrix parameters are presented. In Section 3 the BCS linear classifier is described, pooled estimators of the orthogonally transformed matrix parameters are derived and their distributional properties are established. In Section 4, performance accuracy of the estimated BCS classifier is evaluated using asymptotic expansion of the EPMC. Asymptotics for the moments of the BCS classifier are derived in Section 5. Evaluation of posterior PMC (PPMC) and expected PMC (EPMC) are done in Section 6. Simulation studies are performed for a number of high-dimensional scenarios and advantages of taking structured covariance matrix are numerically verified in Section 7 and finally, conclusions and scope for the future is presented in Section 8.

2 Preliminaries

2.1 Blocked compound symmetry covariance structure

A BCS structure can be written as

$$\begin{aligned}\mathbf{\Gamma} &= \begin{bmatrix} \mathbf{\Sigma}_0 & \mathbf{\Sigma}_1 & \dots & \mathbf{\Sigma}_1 \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \mathbf{\Sigma}_1 & \mathbf{\Sigma}_1 & \dots & \mathbf{\Sigma}_0 \end{bmatrix} \\ &= \mathbf{I}_u \otimes (\mathbf{\Sigma}_0 - \mathbf{\Sigma}_1) + \mathbf{J}_u \otimes \mathbf{\Sigma}_1, \end{aligned} \tag{2.1}$$

where $\mathbf{I}_u$ is the $u \times u$ identity matrix, $\mathbf{1}_u$ is a $u \times 1$ vector of ones, $\mathbf{J}_u = \mathbf{1}_u \mathbf{1}'_u$ and $\otimes$ represents the Kronecker product. We assume $\mathbf{\Sigma}_0$ is a positive definite symmetric $m \times m$ matrix, $\mathbf{\Sigma}_1$ is a symmetric $m \times m$ matrix, under the constraints $-\frac{1}{u-1}\mathbf{\Sigma}_0 < \mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_1 < \mathbf{\Sigma}_0$ (for a proof, see Lemma 2.1 in Roy and Leiva (2011)), so that the $mu \times mu$ matrix $\mathbf{\Gamma}$ is positive definite. Leiva (2007) named $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$ as the equicorrelated matrix parameters. The $m \times m$ block diagonals $\mathbf{\Sigma}_0$ in $\mathbf{\Gamma}$ represent the variance-covariance matrix of the m response variables at any given time, whereas the $m \times m$ block off diagonals $\mathbf{\Sigma}_1$ in $\mathbf{\Gamma}$ represent the covariance matrix of the m response variables between any two time points. We also assume that $\mathbf{\Sigma}_0$ is constant for all time points and $\mathbf{\Sigma}_1$ is constant between any two time points. Our two-level model allows for varying the strength of dependence over time by choosing the covariance matrix $\mathbf{\Sigma}_1$.

2.2 Matrix results

Lemma 4.3 in Ritter and Gallegos (2002) and Leiva (2007) guarantee that a $mu \times mu$ dimensional matrix $\mathbf{\Gamma}$ of the form

$$\mathbf{\Gamma} = \mathbf{I}_u \otimes (\mathbf{\Sigma}_0 - \mathbf{\Sigma}_1) + \mathbf{J}_u \otimes \mathbf{\Sigma}_1,$$

is non singular if both $\mathbf{\Sigma}_0 - \mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_0 + (u-1)\mathbf{\Sigma}_1$ are non singular matrices, and the inverse of $\mathbf{\Gamma}$ is given by

$$\mathbf{\Gamma}^{-1} = \mathbf{I}_u \otimes (\mathbf{\Sigma}_0 - \mathbf{\Sigma}_1)^{-1} + \mathbf{J}_u \otimes \frac{1}{u}\left[(\mathbf{\Sigma}_0 + (u-1)\mathbf{\Sigma}_1)^{-1} - (\mathbf{\Sigma}_0 - \mathbf{\Sigma}_1)^{-1}\right]. \quad (2.2)$$

We notice that $\mathbf{\Gamma}^{-1}$ has the same structure as $\mathbf{\Gamma}$. This result (2.2) generalizes the one given by Bartlett (1951) for $m = 1$.

Using the above results we have the following lemma.

Lemma 1 *Let $\mathbf{\Gamma}$ be a BCS covariance matrix as in equation (2.1). If*

$$\begin{aligned} \mathbf{\Delta}_1 &= \mathbf{\Sigma}_0 - \mathbf{\Sigma}_1, \\ \textit{and} \quad \mathbf{\Delta}_2 &= \mathbf{\Sigma}_0 + (u-1)\,\mathbf{\Sigma}_1 \end{aligned} \quad (2.3)$$

are non singular matrices then $\mathbf{\Gamma}$ is a non singular matrix, and

$$\mathbf{\Gamma}^{-1} = \boldsymbol{I}_u \otimes \mathbf{\Delta}_1^{-1} + \boldsymbol{J}_u \otimes \frac{1}{u}\left(\mathbf{\Delta}_2^{-1} - \mathbf{\Delta}_1^{-1}\right). \quad (2.4)$$

The proof of this lemma is given in Appendix A in Roy and Leiva (2007). This result is used in Section 3 to obtain the BCS linear classifier. At this point we assume that we have samples from p populations Π_p for $p = 1, \ldots, k$ with the common BCS covariance matrix $\mathbf{\Gamma}$. We estimate $\mathbf{\Gamma}$ from each Π_p in the next section.

2.3 Estimation of the equicorrelated parameters $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$ in Π_p

Let $\boldsymbol{x}_{r,s}^{(p)}$ be an m-variate vector of measurements on the r^{th} observation at the s^{th} time point; $r = 1, \ldots, n$, $s = 1, \ldots, u$ from the pth population, $p = 1, \ldots, k$. Let $\boldsymbol{x}_r^{(p)} = (\boldsymbol{x}_{r,1}^{(p)\prime}, \ldots, \boldsymbol{x}_{r,u}^{(p)\prime})'$ be the mu-variate vector of all measurements corresponding to the r^{th} observation. Finally, let $\boldsymbol{x}_1^{(p)}, \boldsymbol{x}_2^{(p)}, \ldots, \boldsymbol{x}_{n^{(p)}}^{(p)}$ be a random sample of size $n^{(p)}$ drawn from the pth population Π_p represented by $N_{mu}\left(\boldsymbol{\nu}^{(p)}, \boldsymbol{\Gamma}\right)$, where $\boldsymbol{\nu}^{(p)} \in \mathbb{R}^{mu}$ and $\boldsymbol{\Gamma}$ is an $mu \times mu$ positive definite BCS structure defined in Section 2.1.

Following Roy and Leiva (2013) unbiased estimates of $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ for Π_p are

$$\widehat{\boldsymbol{\Sigma}}_0^{(p)} = \frac{1}{\left(n^{(p)}-1\right)u} \sum_{r=1}^{n^{(p)}} \sum_{s=1}^{u} \left(\boldsymbol{x}_{r,s}^{(p)} - \widehat{\boldsymbol{\nu}}_{\bullet s}^{(p)}\right)\left(\boldsymbol{x}_{r,s^*}^{(p)} - \widehat{\boldsymbol{\nu}}_{\bullet s^*}^{(p)}\right)',$$

$$\text{and} \quad \widehat{\boldsymbol{\Sigma}}_1^{(p)} = \frac{1}{\left(n^{(p)}-1\right)u\,(u-1)} \sum_{r=1}^{n^{(p)}} \sum_{\substack{s=1 \\ s\neq s^*}}^{u} \sum_{s^*=1}^{u} \left(\boldsymbol{x}_{r,s}^{(p)} - \widehat{\boldsymbol{\nu}}_{\bullet s}^{(p)}\right)\left(\boldsymbol{x}_{r,s^*}^{(p)} - \widehat{\boldsymbol{\nu}}_{\bullet s^*}^{(p)}\right)',$$

respectively. Consequently, an unbiased estimate of $\boldsymbol{\Gamma}$ for the pth population is as follows

$$\widehat{\boldsymbol{\Gamma}}^{(p)} = \boldsymbol{I}_u \otimes \left(\widehat{\boldsymbol{\Sigma}}_0^{(p)} - \widehat{\boldsymbol{\Sigma}}_1^{(p)}\right) + \boldsymbol{J}_u \otimes \widehat{\boldsymbol{\Sigma}}_1^{(p)}, \quad p = 1, \ldots, k.$$

3 Linear classifier under BCS covariance structure

Assuming that Π_p's are represented by $N_{mu}\left(\boldsymbol{\nu}^{(p)}, \boldsymbol{\Gamma}\right)$ we assign a new observation $\boldsymbol{x}$ to $\Pi_{p'}$ whenever $p' = \underset{p=1,\ldots,k}{\arg\max}\, \mathcal{C}\left(\boldsymbol{x}; \boldsymbol{\nu}^{(p)}, \boldsymbol{\Gamma}^{-1}\right)$, where

$$\mathcal{C}\left(\boldsymbol{x}; \boldsymbol{\nu}^{(p)}, \boldsymbol{\Gamma}^{-1}\right) = \boldsymbol{x}' \cdot \boldsymbol{\Gamma}^{-1} \cdot \boldsymbol{\nu}^{(p)} - \frac{1}{2}\boldsymbol{\nu}^{(p)\prime} \cdot \boldsymbol{\Gamma}^{-1} \cdot \boldsymbol{\nu}^{(p)} + \log \pi_p, \tag{3.5}$$

is the linear score for Π_p, $\boldsymbol{\Gamma}^{-1}$ is given in (2.4) and $\sum_{p=1}^{k} \pi_p = 1$ with π_p denoting the prior probability of Π_p.

This classifier is analogous to the well-known Fisher linear discriminant rule that is optimal in a sense of minimum overall misclassification probability (Anderson 2003; Srivastava, 1979).

Evaluation of the performance accuracy of the sample based version of (3.5) requires its distributional properties which in turn depends on the distribution of the sample counterpart $\widehat{\boldsymbol{\Gamma}}^{-1}$ of $\boldsymbol{\Gamma}^{-1}$. A desirable property of the classifier is normality, however it cannot be achieved when the distribution of $\widehat{\boldsymbol{\Gamma}}$ is not Wishart. In this study we circumvent this situation by applying the canonical transformation of the data which is presented in the next section.

3.1 Canonical transformation of the feature data

Let $\boldsymbol{\Xi} = \underset{u\times u}{\boldsymbol{H}'} \otimes \boldsymbol{I}_m$ be an orthogonal matrix with $\boldsymbol{H}$ an orthogonal Helmert matrix whose first column is proportional to a vector of $\mathbf{1}$'s. Making the canonical transformation of the data as,

$$\boldsymbol{y}^{(p)} = \boldsymbol{\Xi}\boldsymbol{x}^{(p)},$$

we have $\boldsymbol{y}^{(p)}$ distributed as $mu-$dimensional normal with

$$\mathrm{E}(\boldsymbol{y}^{(p)}) = \boldsymbol{\mu}^{(p)} = \boldsymbol{\Xi}\mathrm{E}(\boldsymbol{x}^{(p)}) = \boldsymbol{\Xi}\boldsymbol{\nu}^{(p)}, \tag{3.6}$$

and

$$\mathrm{Cov}(\boldsymbol{y}^{(p)}) = \boldsymbol{\Gamma}_{\boldsymbol{\Xi}} = \boldsymbol{\Xi}\boldsymbol{\Gamma}\boldsymbol{\Xi}' = \begin{bmatrix} \boldsymbol{\Delta}_2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{u-1} \otimes \boldsymbol{\Delta}_1 \end{bmatrix}, \tag{3.7}$$

where

$$\begin{aligned} \boldsymbol{\Delta}_1 &= \boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_1, \\ \text{and} \quad \boldsymbol{\Delta}_2 &= \boldsymbol{\Sigma}_0 + (u-1)\,\boldsymbol{\Sigma}_1. \end{aligned}$$

See Lemma 3.1 in Roy and Fonseca (2012) for a proof. Positive definiteness of $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$ is guaranteed as $\boldsymbol{\Gamma}$ is positive definite (see Lemma 2.1 in Roy and Leiva (2011)). We should note that $\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}$ is a block diagonal matrix and $\boldsymbol{\Xi}$ is not a function of either $\boldsymbol{\Sigma}_0$ nor $\boldsymbol{\Sigma}_1$.

Let $T_n = \{\boldsymbol{y}_1^{(p)}, \ldots, \boldsymbol{y}_{n_p}^{(p)}\}_{p\in\{1,\ldots,k\}}$ denote the orthogonally transformed sample data. By noticing the structure of $\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}$ the vectors $\boldsymbol{y}_r^{(p)}$ and $\boldsymbol{\mu}^{(p)}$ are partitioned in u subvectors as $\boldsymbol{y}_r^{(p)} = \left(\boldsymbol{y}_{r,1}^{(p)\prime}, \ldots, \boldsymbol{y}_{r,u}^{(p)\prime}\right)'$ and $\boldsymbol{\mu}^{(p)} = \left(\boldsymbol{\mu}_1^{(p)\prime}, \ldots, \boldsymbol{\mu}_u^{(p)\prime}\right)'$. Then $\boldsymbol{y}_{r,1}^{(p)}, \boldsymbol{y}_{r,2}^{(p)}, \ldots, \boldsymbol{y}_{r,u}^{(p)}$ are independently normally distributed as

$$\begin{aligned} &\boldsymbol{y}_{r,1}^{(p)} \sim N_m\big(\boldsymbol{\mu}_1^{(p)}; \boldsymbol{\Delta}_2\big), \\ \text{and} \quad &\boldsymbol{y}_{r,s}^{(p)} \sim N_m\big(\boldsymbol{\mu}_s^{(p)}; \boldsymbol{\Delta}_1\big) \quad \text{for fixed } s = 2, \ldots, u \quad \text{and fixed } r = 1, \ldots, n. \end{aligned}$$

Now, for each fixed r, we observe that the vector $\boldsymbol{y}_r^{(p)}$ can be partitioned as $\boldsymbol{y}_r^{(p)} = \left(\boldsymbol{y}_{r,1}^{(p)\prime}, \boldsymbol{y}_{r,-1}^{(p)\prime}\right)'$, where we denote $\boldsymbol{y}_{r,-1}^{(p)} = \left(\boldsymbol{y}_{r,2}^{(p)\prime}, \ldots, \boldsymbol{y}_{r,u}^{(p)\prime}\right)'$, i.e., $\boldsymbol{y}_{r,-1}^{(p)}$ is the $u-1$ components of the vector $\boldsymbol{y}_r^{(p)}$ except the first component $\boldsymbol{y}_{r,1}^{(p)}$. Similarly, $\widehat{\boldsymbol{\mu}}^{(p)} = \left(\widehat{\boldsymbol{\mu}}_{\bullet 1}^{(p)\prime}, \ldots, \widehat{\boldsymbol{\mu}}_{\bullet u}^{(p)\prime}\right)'$ with $\widehat{\boldsymbol{\mu}}_{\bullet s}^{(p)} = \frac{1}{n^{(p)}} \sum_{r=1}^{n^{(p)}} \boldsymbol{y}_{r,s}^{(p)}$ for $s = 1, \ldots, u$. We denote $\boldsymbol{\mu}^{(p)} = (\underset{1\times m}{\boldsymbol{\mu}_1^{(p)\prime}}, \underset{1\times(u-1)m}{\boldsymbol{\mu}_{-1}^{(p)\prime}})'$, where $\boldsymbol{\mu}_{-1}^{(p)}$ is the $u-1$ components of the vector $\boldsymbol{\mu}^{(p)}$ except the first component $\boldsymbol{\mu}_1^{(p)}$.

3.2 Estimation of the Transformed Matrix Parameters $\boldsymbol{\Delta}_0$ and $\boldsymbol{\Delta}_1$

From this section onwards we will focus on two population case, i.e., $k = 2$. Now, using the results of Section 2.3, unbiased estimates $\widehat{\boldsymbol{\Delta}}_1^{(p)}$ and $\widehat{\boldsymbol{\Delta}}_2^{(p)}$ of $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$ respectively for the pth population are

$$\begin{aligned} \widehat{\boldsymbol{\Delta}}_1^{(p)} &= \widehat{\boldsymbol{\Sigma}}_0^{(p)} - \widehat{\boldsymbol{\Sigma}}_1^{(p)}, \\ \text{and} \quad \widehat{\boldsymbol{\Delta}}_2^{(p)} &= \widehat{\boldsymbol{\Sigma}}_0^{(p)} + (u-1)\,\widehat{\boldsymbol{\Sigma}}_1^{(p)}, \ \text{ for } \ p = 1, 2. \end{aligned}$$

In the next section we obtain the pooled estimates $\widehat{\boldsymbol{\Delta}}_1$ and $\widehat{\boldsymbol{\Delta}}_2$ of $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$ respectively from both $\widehat{\boldsymbol{\Delta}}_1^{(p)}$ and $\widehat{\boldsymbol{\Delta}}_2^{(p)}$. We also obtain the distributions of $\widehat{\boldsymbol{\Delta}}_1$ and $\widehat{\boldsymbol{\Delta}}_2$.

3.3 Pooled Estimates and Distributions of $\widehat{\boldsymbol{\Delta}}_1$ and $\widehat{\boldsymbol{\Delta}}_2$

Since $\widehat{\boldsymbol{\Gamma}}^{(p)}$ does not follow Wishart distribution, we focus on the distributions of $\widehat{\boldsymbol{\Delta}}_1^{(p)}$ and $\widehat{\boldsymbol{\Delta}}_2^{(p)}$, $p = 1, 2$ which fortunately are both Wishart. Following Theorem 1 in Roy et al. (2013) it can be shown that distributions of $(n^{(p)}-1)(u-1)\widehat{\boldsymbol{\Delta}}_1^{(p)}$ and $(n^{(p)}-1)\widehat{\boldsymbol{\Delta}}_1^{(p)}$ are independent, and

$$\boldsymbol{S}_1^{(p)} = (n^{(p)}-1)(u-1)\widehat{\boldsymbol{\Delta}}_1^{(p)} \quad \sim \quad \text{Wishart}_m\big(\boldsymbol{\Delta}_1, (n^{(p)}-1)(u-1)\big), \tag{3.8a}$$

$$\text{and} \quad \boldsymbol{S}_2^{(p)} = (n^{(p)}-1)\widehat{\boldsymbol{\Delta}}_2^{(p)} \quad \sim \quad \text{Wishart}_m\big(\boldsymbol{\Delta}_2, n^{(p)}-1\big), \tag{3.8b}$$

for $p=1,2$. Now, from the property of Wishart distribution the unbiased estimates of $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$ are $\widehat{\boldsymbol{\Delta}}_1^{(p)}$ and $\widehat{\boldsymbol{\Delta}}_2^{(p)}$ respectively for the pth population. The following lemma yields the pooled estimates of $\widehat{\boldsymbol{\Delta}}_1$ and $\widehat{\boldsymbol{\Delta}}_2$.

Lemma 2 *The pooled estimates of $\widehat{\boldsymbol{\Delta}}_1$ and $\widehat{\boldsymbol{\Delta}}_2$ are given by*

$$\begin{aligned} \widehat{\boldsymbol{\Delta}}_1 &= \frac{\boldsymbol{S}_1}{(n-2)(u-1)}. \\ \textit{and} \quad \widehat{\boldsymbol{\Delta}}_2 &= \frac{\boldsymbol{S}_2}{(n-2)}, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{S}_1 &= \sum_{p=1}^{2}(n^{(p)}-1)(u-1)\widehat{\boldsymbol{\Delta}}_1^{(p)}, \\ \boldsymbol{S}_2 &= \sum_{p=1}^{2}(n^{(p)}-1)\widehat{\boldsymbol{\Delta}}_2^{(p)}, \\ \textit{and} \quad n &= n^{(1)}+n^{(2)}. \end{aligned}$$

Proof: Applying the additive property of two independent Wishart distributions from (3.8a) we get

$$\boldsymbol{S}_1 = \boldsymbol{S}_1^{(1)} + \boldsymbol{S}_1^{(2)} \quad \sim \quad \text{Wishart}_m\big(\boldsymbol{\Delta}_1, (n-2)(u-1)\big),$$

and similarly from (3.8b) we get

$$\boldsymbol{S}_2 = \boldsymbol{S}_2^{(1)} + \boldsymbol{S}_2^{(2)} \quad \sim \quad \text{Wishart}_m\big(\boldsymbol{\Delta}_2, n-2\big),$$

Now, again from (3.8a) we get

$$\boldsymbol{S}_1 = \sum_{p=1}^{2}(n^{(p)}-1)(u-1)\widehat{\boldsymbol{\Delta}}_1^{(p)}.$$

Therefore,

$$\widehat{\boldsymbol{\Delta}}_1 = \frac{\sum_{p=1}^{2}(n^{(p)}-1)(u-1)\widehat{\boldsymbol{\Delta}}_1^{(p)}}{\sum_{p=1}^{2}(n^{(p)}-1)(u-1)} = \frac{\boldsymbol{S}_1}{(n-2)(u-1)}. \tag{3.9}$$

Similarly, from (3.8b) we get

$$\boldsymbol{S}_2 = \sum_{p=1}^{2}(n^{(p)}-1)\widehat{\boldsymbol{\Delta}}_2^{(p)}.$$

Therefore,

$$\widehat{\boldsymbol{\Delta}}_2 = \frac{\sum_{p=1}^{2}(n^{(p)}-1)\widehat{\boldsymbol{\Delta}}_2^{(p)}}{\sum_{p=1}^{2}(n^{(p)}-1)} = \frac{\boldsymbol{S}_2}{(n-2)}. \tag{3.10}$$

Hence, the Lemma follows. This Lemma can easily be extended to more than two populations.

Now, taking inverse of both sides of the equations (3.9) and (3.10) separately we get

$$\widehat{\boldsymbol{\Delta}}_1^{-1} = (n-2)(u-1)\boldsymbol{S}_1^{-1}, \tag{3.11a}$$

$$\text{and} \quad \widehat{\boldsymbol{\Delta}}_2^{-1} = (n-2)\boldsymbol{S}_2^{-1}. \tag{3.11b}$$

Therefore,

$$\mathrm{E}(\widehat{\boldsymbol{\Delta}}_1^{-1}) = (n-2)(u-1)\mathrm{E}(\boldsymbol{S}_1^{-1}), \tag{3.12a}$$

$$\text{and} \quad \mathrm{E}(\widehat{\boldsymbol{\Delta}}_2^{-1}) = (n-2)\mathrm{E}(\boldsymbol{S}_2^{-1}). \tag{3.12b}$$

4 Performance accuracy of BCS classifier

As the goal of classification problem is to derive the decision rule that optimizes some measure of performance accuracy, we consider the PMC as its measure. By partitioning the new observed vector $\boldsymbol{y} = (\underset{1\times m}{\boldsymbol{y}_1'}, \underset{1\times(u-1)m}{\boldsymbol{y}_{-1}'})'$ and by substituting $\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}$ from (3.7) into (3.5), the theoretical classifier (3.5) becomes

$$\begin{aligned} \mathcal{C}(\boldsymbol{y};\boldsymbol{\mu}^{(i)},\boldsymbol{\Delta}_1^{-1},\boldsymbol{\Delta}_2^{-1}) &= \mathcal{C}_1(\boldsymbol{y}_1;\boldsymbol{\mu}^{(i)},\boldsymbol{\Delta}_2^{-1}) + \mathcal{C}_{-1}(\boldsymbol{y}_{-1};\boldsymbol{\mu}^{(i)},\boldsymbol{\Delta}_1^{-1}) \\ &= \Big(\boldsymbol{y}_1 - \frac{1}{2}(\boldsymbol{\mu}_1^{(1)}+\boldsymbol{\mu}_1^{(2)})\Big)'\boldsymbol{\Delta}_2^{-1}(\boldsymbol{\mu}_1^{(1)}-\boldsymbol{\mu}_1^{(2)}) \\ &\quad + \Big(\boldsymbol{y}_{-1} - \frac{1}{2}(\boldsymbol{\mu}_{-1}^{(1)}+\boldsymbol{\mu}_{-1}^{(2)})\Big)'(\boldsymbol{I}_{u-1}\otimes\boldsymbol{\Delta}_1^{-1})(\boldsymbol{\mu}_{-1}^{(1)}-\boldsymbol{\mu}_{-1}^{(2)}). \end{aligned} \tag{4.13}$$

Plugging the estimators of $\boldsymbol{\mu}^{(i)}$ and $\boldsymbol{\Delta}_i$ into (4.13) leads to the sample based classifier

$$\begin{aligned} \mathcal{C}(\boldsymbol{y};\widehat{\boldsymbol{\mu}}^{(i)},\widehat{\boldsymbol{\Delta}}_1^{-1},\widehat{\boldsymbol{\Delta}}_2^{-1}) &= \Big(\boldsymbol{y}_1 - \frac{1}{2}(\widehat{\boldsymbol{\mu}}_1^{(1)}+\widehat{\boldsymbol{\mu}}_1^{(2)})\Big)'\widehat{\boldsymbol{\Delta}}_2^{-1}(\widehat{\boldsymbol{\mu}}_1^{(1)}-\widehat{\boldsymbol{\mu}}_1^{(2)}) \\ &\quad + \Big(\boldsymbol{y}_{-1} - \frac{1}{2}(\widehat{\boldsymbol{\mu}}_{-1}^{(1)}+\widehat{\boldsymbol{\mu}}_{-1}^{(2)})\Big)'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1})(\widehat{\boldsymbol{\mu}}_{-1}^{(1)}-\widehat{\boldsymbol{\mu}}_{-1}^{(2)}). \end{aligned} \tag{4.14}$$

Given that the prior probabilities are the same for both populations, a new observation $\boldsymbol{y}$ is to be assigned to Π_1 whenever $\mathcal{C}(\boldsymbol{y};\boldsymbol{\mu}^{(i)},\boldsymbol{\Delta}_1^{-1},\boldsymbol{\Delta}_2^{-1}) > 0$ or $\mathcal{C}(\boldsymbol{y};\widehat{\boldsymbol{\mu}}^{(i)},\widehat{\boldsymbol{\Delta}}_1^{-1},\widehat{\boldsymbol{\Delta}}_2^{-1}) > 0$. Assume henceforth that $\boldsymbol{y}$ is coming from Π_1, i.e $\boldsymbol{y} \in N(\boldsymbol{\mu}^{(1)},\boldsymbol{\Gamma}_{\boldsymbol{\Xi}})$. The symmetry of our classification rule makes posterior PMC if the mean of $\boldsymbol{y}$ is $\boldsymbol{\mu}^{(1)}$ the same as that under the assumption if mean of $\boldsymbol{y}$ is $\boldsymbol{\mu}^{(2)}$.

Throughout the paper we shall consider the following definitions of PMC:

OPMC : *Optimal or Bayes* PMC of $\mathcal{C}_0(\boldsymbol{y};\boldsymbol{\mu}^{(i)},\boldsymbol{\Delta}_1^{-1},\boldsymbol{\Delta}_2^{-1})$ is defined as

$$\mathcal{E}_O = \Pr\Big(\mathcal{C}_0(\boldsymbol{y};\boldsymbol{\mu}^{(i)},\boldsymbol{\Delta}_1^{-1},\boldsymbol{\Delta}_2^{-1}) \leq 0 | \boldsymbol{y} \in N(\boldsymbol{\mu}^{(1)},\boldsymbol{\Gamma}_{\boldsymbol{\Xi}})\Big).$$

PPMC : *Posterior* PMC of $\mathcal{C}(\boldsymbol{y};\widehat{\boldsymbol{\mu}}^{(i)},\widehat{\boldsymbol{\Delta}}_1^{-1},\widehat{\boldsymbol{\Delta}}_2^{-1})$ is defined for a single training sample T_n of size $n=n^{(1)}+n^{(2)}$ from populations Π_1 and Π_2

$$\mathcal{E}_n^{\mathcal{C}}=\Pr\left(\mathcal{C}(\boldsymbol{y};\widehat{\boldsymbol{\mu}}^{(i)},\widehat{\boldsymbol{\Delta}}_1^{-1},\widehat{\boldsymbol{\Delta}}_2^{-1})\le 0|T_n,\boldsymbol{y}\in N(\boldsymbol{\mu}^{(1)},\boldsymbol{\Gamma}_{\boldsymbol{\Xi}})\right).$$

EPMC : *Expected* PMC of $\mathcal{C}(\boldsymbol{y};\widehat{\boldsymbol{\mu}}^{(i)},\widehat{\boldsymbol{\Delta}}_i^{-1})$ is defined as the average of $\mathcal{E}_n^{\mathcal{C}}$ over arbitrary training samples T_n of size n,

$$\mathrm{E}\mathcal{E}_n^{\mathcal{C}}=E_{T_n}\left[\Pr\left(\mathcal{C}(\boldsymbol{y};\widehat{\boldsymbol{\mu}}^{(i)},\widehat{\boldsymbol{\Delta}}_1^{-1},\widehat{\boldsymbol{\Delta}}_2^{-1})\le 0|T_n,\boldsymbol{y}\in N(\boldsymbol{\mu}^{(1)},\boldsymbol{\Gamma}_{\boldsymbol{\Xi}})\right)\right].$$

It should be noted that for any sample based classifier $\mathcal{C}(\boldsymbol{y};\widehat{\boldsymbol{\mu}}^{(i)},\widehat{\boldsymbol{\Delta}}_1^{-1},\widehat{\boldsymbol{\Delta}}_2^{-1})$ both PPMC and EPMC will always be strictly larger than OPMC.

In what follows we use notations $\boldsymbol{\delta}=(\boldsymbol{\delta}_1',\ldots,\boldsymbol{\delta}_u')'=(\underset{1\times m}{\boldsymbol{\delta}_1'},\underset{1\times(u-1)m}{\boldsymbol{\delta}_{-1}'})'$, with $\boldsymbol{\delta}_1=\boldsymbol{\mu}_1^{(1)}-\boldsymbol{\mu}_1^{(2)}$ and $\boldsymbol{\delta}_{-1}=\boldsymbol{\mu}_{-1}^{(1)}-\boldsymbol{\mu}_{-1}^{(2)}$, and observe that due to the normality of $\boldsymbol{y}$, OPMC can easily be computed as $\mathcal{E}_0=\Phi\left(-\mathcal{D}^2/2\right)$ where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and $\mathcal{D}^2$ is the Mahalanobis (squared) distance between Π_1 and Π_2 which by the structure of $\boldsymbol{\mu}^{(i)}$'s and $\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}$ can represented as

$$\begin{aligned}\mathcal{D}^2 &= (\boldsymbol{\mu}_1^{(1)}-\boldsymbol{\mu}_1^{(2)})'\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}^{-1}(\boldsymbol{\mu}_1^{(1)}-\boldsymbol{\mu}_1^{(2)})\\ &= \boldsymbol{\delta}_1'\boldsymbol{\Delta}_2^{-1}\boldsymbol{\delta}_1+\boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\boldsymbol{\Delta}_1^{-1})\boldsymbol{\delta}_{-1}=\mathcal{D}_1^2+\mathcal{D}_{-1}^2,\end{aligned}$$

with

$$\mathcal{D}_1^2 = \boldsymbol{\delta}_1'\boldsymbol{\Delta}_2^{-1}\boldsymbol{\delta}_1, \tag{4.15a}$$

$$\text{and}\quad \mathcal{D}_{-1}^2 = \boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\boldsymbol{\Delta}_1^{-1})\boldsymbol{\delta}_{-1} = \sum_{k=2}^{u}\boldsymbol{\delta}_k'\boldsymbol{\Delta}_1^{-1}\boldsymbol{\delta}_k=\sum_{k=2}^{u}\mathcal{D}_k^2. \tag{4.15b}$$

Observe that $\mathcal{D}_1^2$ depends on u, which is important for our asymptotic considerations.

The advantage of considering OPMC is that it relates the Mahalanobis distance $\mathcal{D}^2$, i.e., a measure of classification complexity for the true underlying model and the optimal performance accuracy which can be achieved by $\mathcal{C}(\boldsymbol{y})$. By the strict monotonicity of $\Phi(\cdot)$ one can see that $\mathcal{D}^2=-2\Phi^{-1}(\mathcal{E}_O)$, and hence under normality of $\boldsymbol{y}$, $\mathcal{E}_O$ and $\mathcal{D}^2$ provide equivalent information about the classification performance.

5 Asymptotics for the moments of $\mathcal{C}(\boldsymbol{y};\widehat{\boldsymbol{\mu}}^{(i)},\widehat{\boldsymbol{\Delta}}_1^{-1},\widehat{\boldsymbol{\Delta}}_2^{-1})$

To obtain PPMC and EPMC of $\mathcal{C}(\boldsymbol{y};\widehat{\boldsymbol{\mu}}^{(i)},\widehat{\boldsymbol{\Delta}}_1^{-1},\widehat{\boldsymbol{\Delta}}_2^{-1})$, we need to obtain its conditional distribution given the data T_n. We begin by the following

Lemma 3 *Under the model (4.14) and estimators (3.11a) and (3.11b) the PPMC is given by* $\mathcal{E}_n^{\mathcal{C}} = (\Phi(-U/\sqrt{V})|\, T_n)$ *where* U *and* V *are conditional mean and conditional variance of the classifier* $\mathcal{C}(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}^{(i)}, \widehat{\boldsymbol{\Delta}}_1^{-1}, \widehat{\boldsymbol{\Delta}}_2^{-1})$.

Proof: To derive the distribution of $\mathcal{C}(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}^{(i)}, \widehat{\boldsymbol{\Delta}}_1^{-1}, \widehat{\boldsymbol{\Delta}}_2^{-1})$ we first define the following statistics

$$\begin{aligned} U_1 &= \left(\widehat{\boldsymbol{\mu}}_1^{(1)} - \widehat{\boldsymbol{\mu}}_1^{(2)}\right)' \widehat{\boldsymbol{\Delta}}_2^{-1} \left(\widehat{\boldsymbol{\mu}}_1^{(1)} - \boldsymbol{\mu}_1^{(1)}\right) - \frac{1}{2}\left(\widehat{\boldsymbol{\mu}}_1^{(1)} - \widehat{\boldsymbol{\mu}}_1^{(2)}\right)' \widehat{\boldsymbol{\Delta}}_2^{-1} \left(\widehat{\boldsymbol{\mu}}_1^{(1)} - \widehat{\boldsymbol{\mu}}_1^{(2)}\right), \\ U_{-1} &= \left(\widehat{\boldsymbol{\mu}}_{-1}^{(1)} - \widehat{\boldsymbol{\mu}}_{-1}^{(2)}\right)' (\boldsymbol{I}_{u-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1}) \left(\widehat{\boldsymbol{\mu}}_{-1}^{(1)} - \boldsymbol{\mu}_{-1}^{(1)}\right) - \frac{1}{2}\left(\widehat{\boldsymbol{\mu}}_{-1}^{(1)} - \widehat{\boldsymbol{\mu}}_{-1}^{(2)}\right)' (\boldsymbol{I}_{u-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1}) \left(\widehat{\boldsymbol{\mu}}_{-1}^{(1)} - \widehat{\boldsymbol{\mu}}_{-1}^{(2)}\right), \end{aligned}$$

$$\begin{aligned} V_1 &= \left(\widehat{\boldsymbol{\mu}}_1^{(1)} - \widehat{\boldsymbol{\mu}}_1^{(2)}\right)' \widehat{\boldsymbol{\Delta}}_2^{-1} \boldsymbol{\Delta}_2 \widehat{\boldsymbol{\Delta}}_2^{-1} \left(\widehat{\boldsymbol{\mu}}_1^{(1)} - \widehat{\boldsymbol{\mu}}_1^{(2)}\right), \\ V_{-1} &= \left(\widehat{\boldsymbol{\mu}}_{-1}^{(1)} - \widehat{\boldsymbol{\mu}}_{-1}^{(2)}\right)' (\boldsymbol{I}_{u-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1} \boldsymbol{\Delta}_1 \widehat{\boldsymbol{\Delta}}_1^{-1}) \left(\widehat{\boldsymbol{\mu}}_{-1}^{(1)} - \widehat{\boldsymbol{\mu}}_{-1}^{(2)}\right), \end{aligned}$$

$$\begin{aligned} Z_1 &= V_1^{-1/2} \left(\widehat{\boldsymbol{\mu}}_1^{(1)} - \widehat{\boldsymbol{\mu}}_1^{(2)}\right)' \widehat{\boldsymbol{\Delta}}_2^{-1} \left(\boldsymbol{y}_1 - \boldsymbol{\mu}_1^{(1)}\right), \\ \text{and} \quad Z_{-1} &= V_{-1}^{-1/2} \left(\widehat{\boldsymbol{\mu}}_{-1}^{(1)} - \widehat{\boldsymbol{\mu}}_{-1}^{(2)}\right)' \widehat{\boldsymbol{\Delta}}_1^{-1} \left(\boldsymbol{y}_{-1} - \boldsymbol{\mu}_{-1}^{(1)}\right). \end{aligned}$$

Then $\mathcal{C}(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}^{(i)}, \widehat{\boldsymbol{\Delta}}_1^{-1}, \widehat{\boldsymbol{\Delta}}_2^{-1})$ can be written as

$$\mathcal{C}(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}^{(i)}, \widehat{\boldsymbol{\Delta}}_1^{-1}, \widehat{\boldsymbol{\Delta}}_2^{-1}) = \mathcal{C}(\boldsymbol{y}_1; \widehat{\boldsymbol{\mu}}_1^{(i)}, \widehat{\boldsymbol{\Delta}}_2^{-1}) + \mathcal{C}(\boldsymbol{y}_{-1}; \widehat{\boldsymbol{\mu}}_{-1}^{(i)}, \widehat{\boldsymbol{\Delta}}_1^{-1}) = V_1^{1/2} Z_1 - U_1 + V_{-1}^{1/2} Z_{-1} - U_{-1}.$$

Now, by the distributional properties of $\widehat{\boldsymbol{\Delta}}_1^{-1}$ and $\widehat{\boldsymbol{\Delta}}_2^{-1}$ (see Section 3.3) one can see that Z_1 and Z_{-1} are independent and follow the standard normal distributions given T_n (i.e., given $\widehat{\boldsymbol{\mu}}^{(i)}$, $\widehat{\boldsymbol{\Delta}}_1^{-1}$ and $\widehat{\boldsymbol{\Delta}}_2^{-1}$) and assuming that $\boldsymbol{y} \in \Pi_1$. Hence

$$\mathcal{C}(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}^{(i)}, \widehat{\boldsymbol{\Delta}}_1^{-1}, \widehat{\boldsymbol{\Delta}}_2^{-1})|T_n, \boldsymbol{y} \in N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Gamma}_{\boldsymbol{\Xi}}) \sim N_{mu}(-U, V), \tag{5.16}$$

where $U = U_1 + U_{-1}$ and $V = V_1 + V_{-1}$, from which the lemma directly follows.

By (5.16) we found that EPMC is $\mathrm{E}\mathcal{E}_n^{\mathcal{C}} = E_{T_n}\left[\Phi\left(-U/\sqrt{V}\right)\right]$. Both $\mathcal{E}_n^{\mathcal{C}}$ and $\mathrm{E}\mathcal{E}_n^{\mathcal{C}}$ provide *exact* results on the classification performance, however getting the closed-form expressions for them is too demanding. We instead focus on the asymptotic approximation of $\mathrm{E}\mathcal{E}_n^{\mathcal{C}}$ within the following high-dimensional framework:

$$\begin{aligned} & A1: \quad n^{(p)} \to \infty, (p = 1, 2),\ \frac{u}{n} \to c_1 \in (0, \infty), (n^{(1)} + n^{(2)} = n),\ \frac{m}{n-2} \to c_2 \in (0, 1), \\ \text{and} \quad & A2: \quad \mathcal{D}^2 = \mathcal{O}(1). \end{aligned}$$

The condition $A1$ relates the BCS structure to the sample size and for $mu \to \infty$ implies that $q = n - m - 2 \to \infty$. The condition $A2$ is to guarantee that the classification problem does not degenerate, i.e., to ensure that the Mahalanobis distance between the Π_1 and Π_2 is bounded as $0 < d_1 \leq \mathcal{D}^2 \leq d_2 < \infty$. Also, to fulfill A2 we require that $\mathcal{D}_k^2 = \mathcal{O}(\frac{1}{n})$ for $k = 2, \ldots, u$, and $\mathcal{D}_1^2 = \mathcal{O}(\frac{1}{n})$.

Further, we turn to stochastic expansion of the conditional moments of $\mathcal{C}(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}^{(i)}, \widehat{\boldsymbol{\Delta}}_1^{-1}, \boldsymbol{\Delta}_2^{-1})$ under $A1$ and define two random variables $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ as follows

$$\begin{aligned} \boldsymbol{\zeta} &= \frac{1}{\sqrt{n}}\left[n^{(1)}(\widehat{\boldsymbol{\mu}}^{(1)} - \boldsymbol{\mu}^{(1)}) + n^{(2)}(\widehat{\boldsymbol{\mu}}^{(2)} - \boldsymbol{\mu}^{(2)})\right], \\ \text{and} \quad \boldsymbol{\eta} &= \sqrt{\frac{n^{(1)}n^{(2)}}{n}}\left[(\widehat{\boldsymbol{\mu}}^{(1)} - \widehat{\boldsymbol{\mu}}^{(2)}) - (\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^{(2)})\right]. \end{aligned}$$

Since $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ are independent and distributed as $N_{mu}(\boldsymbol{0}, \boldsymbol{\Gamma}_{\boldsymbol{\Xi}})$, using the BCS structure of $\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}$ we can express U and V as follows

$$\begin{aligned} U &= -\frac{1}{2}\left[(\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\delta}_1) + \big(\boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\delta}_{-1}\big)\right] - \frac{n^{(1)} - n^{(2)}}{2n^{(1)}n^{(2)}}\left[\boldsymbol{\eta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1 + \boldsymbol{\eta}_{-1}'(\boldsymbol{I}_{u-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\right] \\ &\quad - \frac{1}{\sqrt{n^{(1)}n^{(2)}}}\left[\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1 + \boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\right] + \frac{1}{\sqrt{n}}\left[\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\zeta}_1 + \boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\zeta}_{-1}\right] \\ &\quad + \frac{1}{\sqrt{n^{(1)}n^{(2)}}}\left[\boldsymbol{\zeta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1 + \boldsymbol{\zeta}_{-1}'(\boldsymbol{I}_{u-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\right], \qquad (5.17) \\ \text{and} \quad V &= \boldsymbol{\delta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\delta} + \frac{n}{n^{(1)}n^{(2)}}\boldsymbol{\eta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\eta} + 2\sqrt{\frac{n}{n^{(1)}n^{(2)}}}\boldsymbol{\delta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\eta}, \qquad (5.18) \end{aligned}$$

where

$$\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1} = \begin{bmatrix} \widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{u-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1} \end{bmatrix}.$$

We now proceed by expanding both U and V stochastically.

5.1 Stochastic expansion of the conditional mean

Following the technique as in Section 3.1 we partition $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ as $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1', \ldots, \boldsymbol{\zeta}_u') = (\underset{1\times m}{\boldsymbol{\zeta}_1'}, \underset{1\times(u-1)m}{\boldsymbol{\zeta}_{-1}'})'$ and $\boldsymbol{\eta}$ as $\boldsymbol{\eta} = (\boldsymbol{\eta}_1', \ldots, \boldsymbol{\eta}_u') = (\underset{1\times m}{\boldsymbol{\eta}_1'}, \underset{1\times(u-1)m}{\boldsymbol{\eta}_{-1}'})'$.

To evaluate the first term in U we need $\mathrm{E}(\widehat{\boldsymbol{\Delta}}_1^{-1})$ and $\mathrm{E}(\widehat{\boldsymbol{\Delta}}_2^{-1})$ that are derived in (3.12a) and (3.12b) respectively. Now using $\mathcal{D}_1^2$ and $\mathcal{D}_{-1}^2$ from (4.15a) and (4.15b) it is seen that the first term in (5.17) have the following representation

$$\begin{aligned} &-\frac{1}{2}\left[\mathrm{E}(\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\delta}_1) + \mathrm{E}\big(\boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1} \otimes \widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\delta}_{-1}\big)\right] \\ &= -\frac{1}{2}\Big[(n-2)\mathrm{E}(\boldsymbol{\delta}_1'\boldsymbol{S}_2^{-1}\boldsymbol{\delta}_1) + (n-2)(u-1)\sum_{k=2}^{u}\mathrm{E}(\boldsymbol{\delta}_k'\boldsymbol{S}_1^{-1}\boldsymbol{\delta}_k) \\ &= -\frac{1}{2}\Big[(n-2)\mathrm{E}(\mathrm{tr}\boldsymbol{W}_2^{-1}\boldsymbol{\xi}_2\boldsymbol{\xi}_2') + \sum_{k-2}^{u}(n-2)\mathrm{E}(\mathrm{tr}\boldsymbol{W}_1^{-1}\boldsymbol{\xi}_{1k}\boldsymbol{\xi}_{1k}')\Big] \\ &= -\frac{(n-2)}{2}\left[\Big(\frac{1}{q} + \frac{1}{q^2}\Big)\mathcal{D}_1^2 + \Big(\frac{1}{q} + \frac{1}{q^2}\Big)\mathcal{D}_{-1}^2\right] + \mathcal{O}(n^{-2}) \\ &= -\frac{(n-2)}{2}\left[\Big(\frac{1}{q} + \frac{1}{q^2}\Big)\mathcal{D}^2\right] + \mathcal{O}(n^{-2}), \qquad (5.19) \end{aligned}$$

where $\boldsymbol{\xi}_2 = \boldsymbol{\Delta}_2^{-1/2}\boldsymbol{\delta}_1$ and $\boldsymbol{\xi}_{1k} = \boldsymbol{\Delta}_1^{-1/2}\boldsymbol{\delta}_k$ for $k = 1, \dots, u$, and

$$\boldsymbol{W}_1 = \frac{1}{(u-1)}\boldsymbol{\Delta}_1^{-1/2}\boldsymbol{S}_1\boldsymbol{\Delta}_1^{-1/2}, \tag{5.20a}$$

$$\text{and} \quad \boldsymbol{W}_2 = \boldsymbol{\Delta}_2^{-1/2}\boldsymbol{S}_2\boldsymbol{\Delta}_2^{-1/2}. \tag{5.20b}$$

Now, using the expressions of $\widehat{\boldsymbol{\Delta}}_1^{-1}$ and $\widehat{\boldsymbol{\Delta}}_2^{-1}$ from (3.11a) and (3.11b), $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ from (5.20a) and (5.20b), and also using the higher order moments of inverse Wishart matrix (see the appendix A.2 in Kubokawa et al. (2013)) we express the expectation of the second term in (5.17) as follows

$$\begin{aligned}
& -\frac{n^{(1)}-n^{(2)}}{2n^{(1)}n^{(2)}}\mathrm{E}\left[\boldsymbol{\eta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1 + \boldsymbol{\eta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\right] \\
= & -\frac{n^{(1)}-n^{(2)}}{2n^{(1)}n^{(2)}}\left[\mathrm{E}\big(\boldsymbol{\eta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1\big) + \mathrm{E}\big(\boldsymbol{\eta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\big)\right] \\
= & -\frac{n^{(1)}-n^{(2)}}{2n^{(1)}n^{(2)}}\left[(n-2)\mathrm{E}\big(\mathrm{tr}(\boldsymbol{\eta}_1'\boldsymbol{S}_2^{-1}\boldsymbol{\eta}_1)\big) + (n-2)(u-1)\mathrm{E}\big(\boldsymbol{\eta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\boldsymbol{S}_1^{-1})\boldsymbol{\eta}_{-1}\big)\right] \\
= & -\frac{n^{(1)}-n^{(2)}}{2n^{(1)}n^{(2)}}\left[(n-2)\mathrm{E}\big(\mathrm{tr}(\boldsymbol{W}_2^{-1})\big) + (n-2)(u-1)\sum_{k=2}^{u}\mathrm{E}\big(\boldsymbol{\eta}_k'\boldsymbol{S}_1^{-1}\boldsymbol{\eta}_k\big)\right] \\
= & -\frac{n^{(1)}-n^{(2)}}{2n^{(1)}n^{(2)}}\left[(n-2)\mathrm{E}\big(\mathrm{tr}(\boldsymbol{W}_2^{-1})\big) + (n-2)\sum_{k=2}^{u}\mathrm{E}\big(\mathrm{tr}(\boldsymbol{W}_1^{-1})\big)\right] \\
= & -\frac{n^{(1)}-n^{(2)}}{2n^{(1)}n^{(2)}}(n-2)mu\left[\frac{1}{q}+\frac{1}{q^2}\right] + \mathcal{O}(n^{-2}).
\end{aligned} \tag{5.21}$$

Using the above expansions (5.19) and (5.21), and using the facts that $\mathrm{E}(\boldsymbol{\zeta}) = 0$ and $\mathrm{E}(\boldsymbol{\eta}) = 0$ we can now express U as $U = \mathscr{U}_0 + \mathscr{U}_1 + \mathscr{U}_2 + \mathcal{O}_P(n^{-2})$ where

$$\mathscr{U}_0 = -\frac{n-2}{2}\left[\frac{1}{q}+\frac{1}{q^2}\right]\left[\mathcal{D}^2 + \frac{n^{(1)}-n^{(2)}}{n^{(1)}n^{(2)}}mu\right],$$

$$\begin{aligned}
\mathscr{U}_1 = & -\frac{\sqrt{n^{(1)}}}{\sqrt{nn^{(2)}}}\Big(\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1 + \boldsymbol{\delta}_{-1}'\big(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\big)\boldsymbol{\eta}_{-1}\Big) + \frac{1}{\sqrt{n}}\left[\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\zeta}_1 + \boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\zeta}_{-1}\right] \\
& + \frac{1}{\sqrt{n^{(1)}n^{(2)}}}\left[\boldsymbol{\zeta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1 + \boldsymbol{\zeta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\right],
\end{aligned}$$

$$\begin{aligned}
\text{and} \quad \mathscr{U}_2 = & -\frac{1}{2}\left[\Big(\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_{-1}'\big(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\big)\boldsymbol{\delta}_{-1}\Big) - (n-2)\Big(\frac{1}{q}+\frac{1}{q^2}\Big)\big(\mathcal{D}_1^2 + \mathcal{D}_{-1}^2(u)\big)\right] \\
& -\frac{n^{(1)}-n^{(2)}}{2n^{(1)}n^{(2)}}\left[\boldsymbol{\eta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2'\Big(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\Big)\boldsymbol{\eta}_2 - (n-2)mu\Big(\frac{1}{q}+\frac{1}{q^2}\Big)\right].
\end{aligned}$$

Remark 1 *By setting $u = 1$, our two-level model reduces to one-level model, consequently $\mathscr{U}_0$ reduces to*

$$\mathscr{U}_0 = -\frac{n-2}{2}\left[\frac{1}{q}+\frac{1}{q^2}\right]\left[\mathcal{D}^2 + \frac{n^{(1)}-n^{(2)}}{n^{(1)}n^{(2)}}m\right].$$

This is the expression of $\mathscr{U}_0$ up to the second order term. Now, by ignoring the second order term in this expression it further reduces to

$$\mathscr{U}_0 = -\frac{n-2}{2q}\left[\mathcal{D}^2 + \frac{n^{(1)}-n^{(2)}}{n^{(1)}n^{(2)}}m\right],$$

which is the expression of U_0 in Kubokawa et al. (2013) (see Page 498).

5.2 Stochastic expansion of the conditional variance

Now we consider the stochastic expansion of V. Considering the first term in (5.18) and substituting the estimators $\widehat{\boldsymbol{\Delta}}_1^{-1}$ and $\widehat{\boldsymbol{\Delta}}_2^{-1}$ from (3.11a) and (3.11b) we get

$$\begin{aligned}\boldsymbol{\delta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\delta} &= \boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_{-1}'\big(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1}\big)\boldsymbol{\delta}_{-1}\\ &= (n-2)^2\boldsymbol{\delta}_1'\boldsymbol{S}_2^{-1}\boldsymbol{\Delta}_2\boldsymbol{S}_2^{-1}\boldsymbol{\delta}_1 + (n-2)^2(u-1)^2\boldsymbol{\delta}_{-1}'\big(\boldsymbol{I}_{u-1}\otimes\boldsymbol{S}_1^{-1}\boldsymbol{\Delta}_1\boldsymbol{S}_1^{-1}\big)\boldsymbol{\delta}_{-1}.\end{aligned}$$

Therefore, using the expressions $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ from (5.20a) and (5.20b), and also using the higher order moments of inverse Wishart matrix (see the appendix A.2 in Kubokawa et al. (2013)) the expectation of the above expression becomes

$$\begin{aligned}&\mathrm{E}\left[(n-2)^2\boldsymbol{\delta}_1'\boldsymbol{S}_2^{-1}\boldsymbol{\Delta}_2\boldsymbol{S}_2^{-1}\boldsymbol{\delta}_1 + (n-2)^2(u-1)^2\boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\boldsymbol{S}_1^{-1}\boldsymbol{\Delta}_1\boldsymbol{S}_1^{-1})\boldsymbol{\delta}_{-1}\right]\\
=\ & (n-2)^2\mathrm{E}[\boldsymbol{\delta}_1'\boldsymbol{S}_2^{-1}\boldsymbol{\Delta}_2\boldsymbol{S}_2^{-1}\boldsymbol{\delta}_1] + (n-2)^2(u-1)^2\sum_{k=2}^{u}\mathrm{E}[\boldsymbol{\delta}_k'\boldsymbol{S}_1^{-1}\boldsymbol{\Delta}_1\boldsymbol{S}_1^{-1}\boldsymbol{\delta}_k]\\
=\ & (n-2)^2\mathrm{E}[\mathrm{tr}\boldsymbol{W}_2^{-2}\boldsymbol{\xi}_2\boldsymbol{\xi}_2'] + (n-2)^2\sum_{k=2}^{u}\mathrm{E}[\mathrm{tr}\boldsymbol{W}_1^{-2}\boldsymbol{\xi}_{1k}\boldsymbol{\xi}_{1k}']\\
=\ & \frac{(n-2)^3}{q^3}\mathcal{D}_1^2 + \frac{(n-2)^2\big(4(n-2)-q\big)}{q^4}\mathcal{D}_1^2\\
& + \sum_{k=2}^{u}\left[\frac{(n-2)^3}{q^3}\mathcal{D}_k^2 + \frac{(n-2)^2\big(4(n-2)-q\big)}{q^4}\mathcal{D}_k^2\right] + \mathcal{O}(n^{-2})\\
=\ & \left[\frac{(n-2)^3}{q^3} + \frac{(n-2)^2\big(4(n-2)-q\big)}{q^4}\right]\mathcal{D}^2 + \mathcal{O}(n^{-2}). \qquad (5.22)\end{aligned}$$

Now, considering the second term in (5.18) and substituting the values of $\widehat{\boldsymbol{\Delta}}_1^{-1}$ and $\widehat{\boldsymbol{\Delta}}_2^{-1}$ from (3.11a) and (3.11b) we get

$$\begin{aligned}\frac{n}{n^{(1)}n^{(2)}}\boldsymbol{\eta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\eta} &= \frac{n}{n^{(1)}n^{(2)}}\left[\boldsymbol{\eta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1 + \boldsymbol{\eta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\right]\\ &= \frac{n}{n^{(1)}n^{(2)}}\left[\boldsymbol{\eta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1 + \boldsymbol{\eta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\right].\end{aligned}$$

Now,

$$\begin{aligned}&\mathrm{E}\left[\Big(\boldsymbol{\eta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1 + \boldsymbol{\eta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\Big)\right]\\
=\ & \mathrm{E}\left[\boldsymbol{\eta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1\right] + \mathrm{E}\left[\boldsymbol{\eta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\right]\\
=\ & (n-2)^2\mathrm{E}\left[\boldsymbol{\eta}_1'\boldsymbol{S}_2^{-1}\boldsymbol{\Delta}_2\boldsymbol{S}_2^{-1}\boldsymbol{\eta}_1\right] + (n-2)^2(u-1)^2\mathrm{E}\left[\boldsymbol{\eta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\boldsymbol{S}_1^{-1}\boldsymbol{\Delta}_1\boldsymbol{S}_1^{-1})\boldsymbol{\eta}_{-1}\right]\\
=\ & (n-2)^2\mathrm{E}\left[\boldsymbol{\eta}_1'\boldsymbol{S}_2^{-1}\boldsymbol{\Delta}_2\boldsymbol{S}_2^{-1}\boldsymbol{\eta}_1\right] + (u-1)^2\sum_{k=2}^{u}(n-2)^2\mathrm{E}\left[\boldsymbol{\eta}_k'(\boldsymbol{S}_1^{-1}\boldsymbol{\Delta}_1\boldsymbol{S}_1^{-1})\boldsymbol{\eta}_k\right]\\
=\ & (n-2)^2\mathrm{E}\left[\mathrm{tr}\boldsymbol{W}_2^{-2}\right] + \sum_{k=2}^{u}(n-2)^2\mathrm{E}\left[\mathrm{tr}\boldsymbol{W}_1^{-2}\right]\\
=\ & \left[\frac{m(n-2)^2}{q^3} + \frac{m(n-2)\big(4(n-2)-q\big)}{q^4}\right](n-2)u + \mathcal{O}(n^{-2}).\end{aligned}$$

Therefore, the expectation of the second term in (5.18) is as follows

$$\frac{n}{n^{(1)}n^{(2)}}\left[\left(\frac{m(n-2)^2}{q^3}+\frac{m(n-2)\big(4(n-2)-q\big)}{q^4}\right)(n-2)u\right]+\mathcal{O}(n^{-2}). \tag{5.23}$$

Therefore, using the above expansions (5.22) and (5.23) we can now express V as $V=\mathscr{V}_0+\mathscr{V}_1+\mathscr{V}_2+\mathcal{O}_P(n^{-3/2})$ where

$$\begin{aligned}
\mathscr{V}_0 &= \left[\frac{(n-2)^3}{q^3}+\frac{(n-2)^2\big(4(n-2)-q\big)}{q^4}\right]\Big[\mathcal{D}^2+\frac{n}{n^{(1)}n^{(2)}}mu\Big],\\
\mathscr{V}_1 &= 2\sqrt{\frac{n}{n^{(1)}n^{(2)}}}\left[\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1+\boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\right],\\
\text{and}\quad \mathscr{V}_2 &= \boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\delta}_1+\boldsymbol{\delta}_{-1}'\big(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1}\big)\boldsymbol{\delta}_{-1}-\left[\frac{(n-2)^3}{q^3}+\frac{(n-2)^2\big(4(n-2)-q\big)}{q^4}\right]\mathcal{D}^2\\
&\quad+\frac{n}{n^{(1)}n^{(2)}}\left[\boldsymbol{\eta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1+\boldsymbol{\eta}_{-1}'\big(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1}\big)\boldsymbol{\eta}_{-1}\right]\\
&\quad-\frac{n}{n^{(1)}n^{(2)}}\left[\left(\frac{(n-2)^3}{q^3}+\frac{(n-2)^2\big(4(n-2)-q\big)}{q^4}\right)mu\right].
\end{aligned}$$

Remark 2 *As in Remark 1, by setting $u=1$, the model reduces to one-level model, consequently $\mathscr{V}_0$ reduces to*

$$\mathscr{V}_0 = \left[\frac{(n-2)^3}{q^3}+\frac{(n-2)^2\big(4(n-2)-q\big)}{q^4}\right]\Big[\mathcal{D}^2+\frac{n}{n^{(1)}n^{(2)}}m\Big].$$

This is the expression of $\mathscr{V}_0$ up to the second order term. Now, by ignoring the second order term in this expression it further reduces to

$$\mathscr{V}_0 = \left[\frac{(n-2)^3}{q^3}\right]\Big[\mathcal{D}^2+\frac{n}{n^{(1)}n^{(2)}}m\Big],$$

which is the expression of V_0 in Kubokawa et al. (2013) (see Page 499).

6 Evaluation of PPMC and EPMC

To evaluate PPMC we first obtain expansion of the distribution function of $\mathcal{C}(\boldsymbol{y};\widehat{\boldsymbol{\mu}}^{(i)},\widehat{\boldsymbol{\Delta}}_1^{-1},\widehat{\boldsymbol{\Delta}}_2^{-1})$. By returning now to Lemma 3 and (5.16), and by using the representations $U=\mathscr{U}_0+\mathscr{U}_1+\mathscr{U}_2+\mathcal{O}_P(n^{-3/2})$ and $V=\mathscr{V}_0+\mathscr{V}_1+\mathscr{V}_2+\mathcal{O}_P(n^{-3/2})$ we consider the following Taylor series expansion

$$\begin{aligned}
\frac{U}{V^{1/2}} &= \frac{\mathscr{U}_0+\mathscr{U}_1+\mathscr{U}_2}{\mathscr{V}_0^{1/2}}\left(1+\frac{\mathscr{V}_1+\mathscr{V}_2}{\mathscr{V}_0}\right)^{-1/2}\\
&= \frac{(\mathscr{U}_0+\mathscr{U}_1+\mathscr{U}_2)}{\mathscr{V}_0^{1/2}}-\frac{(\mathscr{U}_0+\mathscr{U}_1+\mathscr{U}_2)}{\mathscr{V}_0^{1/2}}\frac{\mathscr{V}_1}{2\mathscr{V}_0}-\frac{(\mathscr{U}_0+\mathscr{U}_1+\mathscr{U}_2)}{\mathscr{V}_0^{1/2}}\frac{\mathscr{V}_2}{2\mathscr{V}_0}\\
&\quad+\frac{3}{8}\frac{(\mathscr{U}_0+\mathscr{U}_1+\mathscr{U}_2)}{\mathscr{V}_0^{1/2}}\frac{\mathscr{V}_1^2}{\mathscr{V}_0^2}+\frac{3}{8}\frac{(\mathscr{U}_0+\mathscr{U}_1+\mathscr{U}_2)}{\mathscr{V}_0^{1/2}}\frac{\mathscr{V}_2^2}{\mathscr{V}_0^2}+\frac{6}{8}\frac{(\mathscr{U}_0+\mathscr{U}_1+\mathscr{U}_2)}{\mathscr{V}_0^{1/2}}\frac{\mathscr{V}_1\mathscr{V}_2}{\mathscr{V}_0^2}+\mathcal{O}_P(n^{-3/2}).
\end{aligned}$$

Further, by ignoring the terms of order higher than $\mathcal{O}_P(n^{-3/2})$ we obtain

$$\frac{U}{V^{1/2}} = \gamma_0 + \gamma_1 + \gamma_2 + \mathcal{O}_P(n^{-3/2}),$$

$$\begin{aligned} \text{where} \quad \gamma_0 &= \frac{\mathscr{U}_0}{\mathscr{V}_0^{1/2}}, \\ \gamma_1 &= \mathscr{V}_0^{-1/2}\Big[\mathscr{U}_1 - \frac{\mathscr{U}_0}{2\mathscr{V}_0}\mathscr{V}_1\Big], \\ \text{and} \quad \gamma_2 &= \mathscr{V}_0^{-1/2}\Big[\mathscr{U}_2 - \frac{\mathscr{U}_0}{2\mathscr{V}_0}\mathscr{V}_2 + \frac{3\mathscr{U}_0}{8\mathscr{V}_0^2}\mathscr{V}_1^2 - \frac{1}{2\mathscr{V}_0}\mathscr{U}_1\mathscr{V}_1\Big]. \end{aligned}$$

The above results can be summarized in the following theorem.

Theorem 1 *The cumulative distribution function of the normalized classifier* $\Big[\mathcal{C}(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}^{(i)}, \widehat{\boldsymbol{\Delta}}_1^{-1}, \widehat{\boldsymbol{\Delta}}_2^{-1}) - U)\Big] / V^{1/2}$ *under* $\boldsymbol{y} \in \Pi_1$ *and under the condition A*1 *is expanded as*

$$\Phi\big(\gamma_0 + \gamma_1 + \gamma_2\big) + \mathcal{O}_P(n^{-3/2}) \tag{6.24}$$

where $\phi(\cdot)$ is the density function of $N(0,1)$.

Now, by considering the Taylor series expansion of $\Phi\big(\gamma_0 + \gamma_1 + \gamma_2\big)$ in (6.24) around γ_0 and using the fact that $\phi'(x) = -x\phi(x)$ one can see that PPMC of $\mathcal{C}(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}^{(i)}, \widehat{\boldsymbol{\Delta}}_1^{-1}, \widehat{\boldsymbol{\Delta}}_2^{-1})$ is expanded as follows

$$\mathcal{E}_n^{\mathcal{C}} = \Phi(\gamma_0) + \phi(\gamma_0)\Big(\gamma_1 + \gamma_2 - \frac{1}{2}\gamma_0\gamma_1^2\Big) + \mathcal{O}_P(n^{-3/2}). \tag{6.25}$$

Remark 3 *For* $u = 1$, $\Phi(\gamma_0)$ *reduces to*

$$\Phi(\gamma_0) \quad = \quad \Phi\left(\frac{\mathscr{U}_0}{\mathscr{V}_0^{1/2}}\right),$$

which represents the corresponding expression in Kubokawa et al. (2013) (see Page 500) up to the second order term.

Thus, from Remarks 1, 2 and 3 we see that our method is a clear extension of Kubokawa et al. (2013) for the two-level multivariate observations.

To evaluate EPMC, we denote $\mathscr{L} = \mathrm{E}\left(\gamma_1 + \gamma_2 - \frac{1}{2}\gamma_0\gamma_1^2\right)$ and observe that $\mathscr{L}$ can be written as

$$\begin{aligned} \mathscr{L} \quad = \quad & \mathscr{V}_0^{-1/2}\Big[\mathrm{E}(\mathscr{U}_1) + \mathrm{E}(\mathscr{U}_2)\Big] - \frac{\mathscr{U}_0}{2\mathscr{V}_0^{3/2}}\Big[\mathrm{E}(\mathscr{V}_1) + \mathrm{E}(\mathscr{V}_2)\Big] - \frac{\mathscr{U}_0}{2\mathscr{V}_0^{3/2}}\mathrm{E}(\mathscr{U}_1^2) \\ & + \frac{\mathscr{U}_0}{8\mathscr{V}_0^{5/2}}\left(3 - \frac{\mathscr{U}_0^2}{\mathscr{V}_0}\right)\mathrm{E}(\mathscr{V}_1^2) - \frac{1}{2\mathscr{V}_0^{3/2}}\left(1 - \frac{\mathscr{U}_0^2}{\mathscr{V}_0}\right)\mathrm{E}(\mathscr{U}_1\mathscr{V}_1). \end{aligned}$$

Now, to calculate the approximate EPMC, we need to evaluate the expectation of each term in $\mathscr{L}$.

By using the fact that $\boldsymbol{\zeta}$ and $\boldsymbol{\eta}$ are independent and identically distributed while evaluation $\mathrm{E}(\mathscr{U}_1)$, and by using the expressions (5.19) and (5.21) and by keeping the second order term while evaluation $\mathrm{E}(\mathscr{U}_1)$ respectively, we obtain

$$\begin{aligned} \mathrm{E}(\mathscr{U}_1) &= 0, \\ \text{and} \quad \mathrm{E}(\mathscr{U}_2) &= -\frac{(n-2)}{2(1+q_2)^2}\left[\mathcal{D}^2 + \frac{n^{(1)}-n^{(2)}}{n^{(1)}n^{(2)}}mu\right] + \mathcal{O}(n^{-3/2}). \end{aligned}$$

Again, by applying the similar arguments for $\mathscr{V}_1$ and $\mathscr{V}_2$ we obtain

$$\begin{aligned} \mathrm{E}(\mathscr{V}_1) &= 0, \\ \text{and} \quad \mathrm{E}(\mathscr{V}_2) &= \left(\frac{(n-2)^2\big(4(n-2)-q\big)}{(q)^4}\right)\Big(\mathcal{D}^2 + \frac{nmu}{n^{(1)}n^{(2)}}\Big) + \mathcal{O}(n^{-3/2}). \end{aligned}$$

Now, using the results on the higher order moments of the inverse of Wishart distribution in Appendix A2 in the Kubokawa et al. (2013) we evaluate $\mathrm{E}(\mathscr{V}_1^2)$ as follows.

$$\begin{aligned} \mathrm{E}(\mathscr{V}_1^2) &= \mathrm{E}\Big[2\sqrt{\frac{n}{n^{(1)}n^{(2)}}}\,\Big(\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1 + \boldsymbol{\delta}_{-1}'\big(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1}\big)\boldsymbol{\eta}_{-1}\Big)\Big]^2 \\ &= 4\frac{n}{n^{(1)}n^{(2)}}\Big[\mathrm{E}\Big(\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1\boldsymbol{\eta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\delta}_1 + \sum_{k=2}^{u}\boldsymbol{\delta}_k'\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\eta}_k\boldsymbol{\eta}_k'\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\delta}_k\Big)\Big] \\ &= 4\frac{n(n-2)}{n^{(1)}n^{(2)}}\Big[\frac{(n-2)^4\big\{(n-2)m + ((n-2)+m)^2\big\}}{q^7}\Big]\mathcal{D}^2 + \mathcal{O}(n^{-3/2}). \end{aligned}$$

Further, given that the expectation of the product terms are zeros we now obtain $\mathrm{E}(\mathscr{U}_1^2)$ as follows

$$\begin{aligned} \mathrm{E}(\mathscr{U}_1^2) &= \mathrm{E}\Big[\sqrt{\frac{n^{(1)}}{nn^{(2)}}}\boldsymbol{\delta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\eta} + \frac{1}{\sqrt{n}}\boldsymbol{\delta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\zeta} + \frac{1}{\sqrt{n^{(1)}n^{(2)}}}\boldsymbol{\zeta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\eta}\Big]^2 \\ &= \frac{n^{(1)}}{nn^{(2)}}\mathrm{E}(\boldsymbol{\delta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\eta}\boldsymbol{\eta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\delta}) + \frac{1}{n}\mathrm{E}(\boldsymbol{\delta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\zeta}\boldsymbol{\zeta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\delta}) + \frac{1}{n^{(1)}n^{(2)}}\mathrm{E}(\boldsymbol{\zeta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\eta}\boldsymbol{\eta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\zeta}) \\ &= (\frac{n^{(1)}}{nn^{(2)}} + \frac{1}{n})\mathrm{E}(\boldsymbol{\delta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\delta}) + \frac{1}{n^{(1)}n^{(2)}}\mathrm{E}(\boldsymbol{\zeta}'\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\Gamma}_{\boldsymbol{\Xi}}\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{\Xi}}^{-1}\boldsymbol{\zeta}) \\ &= (\frac{n^{(1)}}{nn^{(2)}} + \frac{1}{n})\mathrm{E}(\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\delta}_{-1})) \\ &\quad + \frac{1}{n^{(1)}n^{(2)}}\mathrm{E}(\boldsymbol{\zeta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\zeta}_1 + \boldsymbol{\zeta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\zeta}_{-1})) \\ &= \left(\frac{n^{(1)}}{nn^{(2)}} + \frac{1}{n}\right)\left[(n-2)^2\mathrm{E}\big(\mathrm{tr}\boldsymbol{W}_2^{-2}\boldsymbol{\xi}_2\boldsymbol{\xi}_2'\big) + (n-2)^2\sum_{k=2}^{u}\mathrm{E}\big(\mathrm{tr}\boldsymbol{W}_1^{-2}\boldsymbol{\xi}_{1k}\boldsymbol{\xi}_{1k}'\big)\right] \\ &\quad + \frac{1}{n^{(1)}n^{(2)}}\left[(n-2)^2\mathrm{E}\big(\mathrm{tr}\boldsymbol{W}_2^{-2}\big) + (n-2)^2\sum_{k=2}^{u}\mathrm{E}\big(\mathrm{tr}\boldsymbol{W}_1^{-2}\big)\right] \\ &= \left[\frac{(n-2)^3}{q^3} + \frac{(n-2)^2\big(4(n-2)-q\big)}{q^4}\right]\left[\mathcal{D}^2\left(\frac{n^{(1)}}{nn^{(2)}} + \frac{1}{n}\right) + \frac{mu}{n^{(1)}n^{(2)}}\right] + \mathcal{O}(n^{-3/2}). \end{aligned}$$

The above results are derived using the independence of $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_{-1}$, independence of $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$, as well as $\mathrm{E}(\boldsymbol{\zeta}) = 0$. Finally, we express $\mathrm{E}(\mathscr{U}_1\mathscr{V}_1)$ as follows

$$\begin{aligned}
\mathrm{E}(\mathscr{U}_1\mathscr{V}_1) &= -\frac{2}{n^{(2)}}\mathrm{E}\left[\left(\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1 + \boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\right)\left(\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1 + \boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\right)\right]\\
&= -\frac{2}{n^{(2)}}\left[\mathrm{E}\left(\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\eta}_1\boldsymbol{\eta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\delta}_1\right) + \mathrm{E}\left(\boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\eta}_{-1}\boldsymbol{\eta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\delta}_{-1}\right)\right]\\
&= -\frac{2}{n^{(2)}}\left[\mathrm{E}\left(\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\delta}_1\right) + \mathrm{E}\left(\boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1})(\boldsymbol{I}_{u-1}\otimes\boldsymbol{\Delta}_1)(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\delta}_{-1}\right)\right]\\
&= -\frac{2}{n^{(2)}}\left[\mathrm{E}\left(\boldsymbol{\delta}_1'\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\Delta}_2\widehat{\boldsymbol{\Delta}}_2^{-1}\boldsymbol{\delta}_1\right) + \mathrm{E}\left(\boldsymbol{\delta}_{-1}'(\boldsymbol{I}_{u-1}\otimes\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1}\boldsymbol{\Delta}_1\widehat{\boldsymbol{\Delta}}_1^{-1})\boldsymbol{\delta}_{-1}\right)\right]\\
&= -\frac{2}{n^{(2)}}\left[(n-2)^4\frac{(n-2+m)}{q^5}\mathcal{D}_1^2 + (n-2)^4\frac{(n-2+m)}{q^5}\sum_{k=2}^{u}\mathcal{D}_k^2\right] + \mathcal{O}(n^{-3/2})\\
&= -\frac{2}{n^{(2)}}(n-2)^4\frac{(n-2+m)}{q^5}\mathcal{D}^2 + \mathcal{O}(n^{-3/2}).
\end{aligned}$$

Now, by summarizing the above results we obtain the following theorem.

Theorem 2 *Under the conditions A1 and A2 the asymptotic approximation of EPMC of* $\mathcal{C}(\boldsymbol{y};\widehat{\boldsymbol{\mu}}^{(i)},\widehat{\boldsymbol{\Delta}}_1^{-1},\widehat{\boldsymbol{\Delta}}_2^{-1})$ *is given by* $E_{T_n}\mathcal{E}_n^{\mathcal{C}} = \Phi(\gamma_0) + \phi(\gamma_0)\mathscr{L}(\mathcal{D}^2)$.

6.1 Relation to existing results

To compare our approximation with the existing results we apply the asymptotic approximations of the type considered in Wyman et al. (1990) and Fujikoshi et al. (2010) to the BCS classifier. As we are mostly interested in the asymptotic effect of u we focus on $\mathcal{C}_{-1} = \mathcal{C}(\boldsymbol{y}_{-1};\widehat{\boldsymbol{\mu}}_{-1}^{(i)},\widehat{\Delta}_1^{-1})$. We recall that by the bounding condition $A2$, we assume that $\mathcal{D}_k^2 = \mathcal{O}(1/n)$ for all $k = 2,\ldots,u$ and observe that with $u\to\infty$, the distribution of $\mathcal{C}_{-1}$ as a sum of growing number of independent and identically distributed random variables converges to the normal distribution, so that EPMC can be approximated as

$$\mathrm{E}_{T_n}\mathcal{E}_n^{\mathcal{C}_{-1}} \approx \Phi\left(\mathrm{E}(\mathcal{U})\left[E(\mathcal{V})\right]^{-1/2}\right),$$

where

$$\mathcal{U} = \sum_{k=2}^{u}\left[\widehat{\boldsymbol{\delta}}_k'\widehat{\Delta}_1^{-1}\left(\boldsymbol{y}_k^0 - \widehat{\boldsymbol{\mu}}_k^{(1)}\right)\right] - \frac{1}{2}\sum_{k=2}^{u}\mathcal{D}_k^2 \quad\text{and}\quad \mathcal{V} = \sum_{k=2}^{u}\widehat{\boldsymbol{\delta}}_k'\widehat{\Delta}_1^{-1}\Delta_1\widehat{\Delta}_1^{-1}\widehat{\boldsymbol{\delta}}_k. \tag{6.26}$$

Now, by evaluating the moments in (6.26) and by deriving the representation of $\mathrm{E}_{T_n}\mathcal{E}_n^{\mathcal{C}_{-1}}$, we obtain

$$\mathrm{E}_{T_n}\mathcal{E}_n^{\mathcal{C}_{-1}} \approx \Phi\left(-\frac{1}{2}\sqrt{\frac{2n^{(0)}}{2n^{(0)}-m}}\frac{\mathcal{D}_{-1}^2(u)}{\sqrt{[\mathcal{D}_{-1}^2(u) + \frac{2mu}{n^{(0)}}]}}\right) = \Phi\left(-\frac{1}{2}\frac{\mathcal{D}_{-1}(u)}{\sqrt{\left[1+\frac{4mu}{n^{(0)}\mathcal{D}_{-1}^2(u)}\right]\left[1+\frac{m}{2n^{(0)}-m}\right]}}\right), \tag{6.27}$$

where we set $n^{(1)} = n^{(2)} = n^{(0)}$ and ignore the terms of order $1/n^{(0)}$ and $m/(n^{(0)})^2$ in order to single out effects of m and u. The second term in the denominator of (6.27) reflects increase in the misclassification rate due to estimation of Δ_1^{-1}: for fixed u, the effect of the block size, m when $n^{(0)}$ grows. However, the effect of u is much more pronounced since $\frac{4mu}{n^{(0)}}\mathcal{D}^2_{-1}(u)$ grows with u thereby increasing $\mathrm{E}_{T_n}\mathcal{E}_n^{\mathcal{C}_{-1}}$. However, the finite sample size behavior of the approximation is much more subtle because it depends on the terms of order $1/n^{(0)}$ and $m/(n^{(0)})^2$.

7 Simulation Studies

We now examine the performance accuracy of the linear classifiers (3.5) with the help of simulation studies. Populations Π_1 and Π_2 are represented by normal distribution with the common BCS covariance matrix $\mathbf{\Gamma}$. The matrix parameters $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$ in $\mathbf{\Gamma}$ are chosen as $m \times m$ identity matrix and $m \times m$ zero matrix with all elements as zeros respectively; and $\mathbf{\Gamma}^{-1/2}\boldsymbol{\nu}_1 = (mu)^{-1/2}(\mathcal{D}, \mathcal{D}, \ldots, \mathcal{D})'$ for $\mathcal{D}^2 = 2, 4$ and $\boldsymbol{\nu}_2 = (0, 0, \ldots, 0)'$. The transformed means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and the transformed variance $\mathbf{\Gamma}_{\mathbf{\Xi}}$ are obtained by the orthogonalization of $\mathbf{\Xi}$, with $u \times u$ dimensional Helmert matrix $\boldsymbol{H}$ as described in Section 3.1. By setting $\mathcal{D}^2 = 2$ and $\mathcal{D}^2 = 4$, we consider two types of classification complexity for two-level data, for which the OMCP $\mathcal{E}_O = 0.223$ and $\mathcal{E}_O = 0.159$ respectively.

To see the effect of the repeated measurements over time u, the values of u are chosen as 3, 5 and 8. Similarly to see the effect of the number of variables m, it is chosen as 2, 5, 8 and 10. The total sample size n is chosen as 20, 40 and 80, with different combinations of the pairs of sample sizes $(n^{(1)}, n^{(2)})$, to encompass both balanced and unbalanced cases. Even though we have chosen $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$ as $m \times m$ identity matrix and $m \times m$ zero matrix, the calculated values of $\Phi(\gamma_0)$, the limiting behavior of EPMC do not depend on the choices of $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}_1$.

Simulation results are given in Tables 1, 2 and 3 for $(n^{(1)} = n^{(2)})$, $(n^{(1)} > n^{(2)})$ and $(n^{(1)} < n^{(2)})$ respectively. The limiting values of EPMC, $\Phi(\gamma_0)$ are given in these tables in the first row and the approximate EPMC values are given in the second row in both *italics* and parenthesis for each combination of $\mathcal{D}^2$, m, u and $(n^{(1)}, n^{(2)})$. Table 4 present the values of $\Phi(\gamma_0)$ for different $\mathcal{D}^2$, m and $(n^{(1)}, n^{(2)})$, but for fixed $u = 1$. As mentioned in the various remarks in the previous sections that the our method reduces to the method of Kubokawa et al. (2013) for $u = 1$, and indeed we get the exact identical values of $\Phi(\gamma_0)$ up to the first order term as Kubokawa et al. (2013) for different values of $\mathcal{D}^2$, m and $(n^{(1)}, n^{(2)})$. Nevertheless, for comparison purpose with our new extended method we also calculate the limiting values of EPMC up to the second order term for $u = 1$.

In Tables 1, 2 and 3 we see that both $\Phi(\gamma_0)$ and the approximate EPMC values increase with m and u

for each fixed $\mathcal{D}^2$ (this result is supported by (6.27)) and $(n^{(1)}, n^{(2)})$. However, both $\Phi(\gamma_0)$ and approximate EPMC values decrease with n (both balanced and unbalanced cases) for each fixed $\mathcal{D}^2$, m and u. This is naturally expected as increasing sample size should improve performance accuracy. It is very important to point out that with our technique both the limiting value $\Phi(\gamma_0)$ and approximate EPMC values work well even for small n, however Kubokawa et al. (2013) being not designed for small sample case fails. Furthermore, for those sample sizes when Kubokawa et al. (2013) works, our proposed method for two-level data outperforms their results. To demonstrate the achieved gain consider for example, two following set-ups, $\mathcal{D}^2 = 2$, $m = 2$, $u = 5$, $(n^{(1)}, n^{(2)}) = (15, 25)$, and $\mathcal{D}^2 = 2$, $m = 5$, $u = 2$, $(n^{(1)}, n^{(2)}) = (15, 25)$. Our method gives $\Phi(\gamma_0) = 0.317$ and $\Phi(\gamma_0) = 0.325$ respectively, whereas the corresponding value in Kubokawa et al. (2013) is 0.338. For additional comparison of the performance pattern of the two methods consider the case when $\mathcal{D}^2 = 2$; with $m = 10$, $u = 5$, $(n^{(1)}, n^{(2)}) = (30, 50)$, in Table 3, and with $m = 50$, $(n^{(1)}, n^{(2)}) = (30, 50)$ in Table 4, which result in 0.387 and 0.429 respectively. Observe that both limiting values are calculated using second order term, thereby showing the gain of taking into account the covariance structure underlying the data.

Another important observation here is that with fixed total dimensionality, larger number of replicates (u) results in improved performance accuracy. For example, given that $mu = 10$, we consider the two cases, $m = 2$, $u = 5$, and $m = 5$ and $u = 2$; which result in 0.317 and 0.325 respectively. See Table 3, for $(n^{(1)}, n^{(2)}) = (15, 25)$ and $\mathcal{D}^2 = 2$. This is in accordance with the theory of repeated measurements where increasing the number of replicates improves precision.

Moreover, we observe another very important fact that when $n^{(1)} > n^{(2)}$ both $\Phi(\gamma_0)$, the limiting values of EPMC are less that their counterparts when $n^{(1)} < n^{(2)}$ and $n^{(1)} = n^{(2)}$. Thus, in practical applications it is suggested that when classifying a patient with a rare disease, as a sick or healthy, one needs to choose diseased Population as Π_2.

It is also interesting to observe the effect of increasing sample size for both choices of the classification complexity ($\mathcal{D}^2 = 2$ and $\mathcal{D}^2 = 4$); the limiting values $\phi(\gamma_0)$ and EPMC becomes closer to each other, see Tables 1, 2 and 3. Observe also that the difference between the first and second order representation of Φ_0 is negligible, see Table 4.

8 Conclusions and scope for the future

We have analyzed the performance accuracy of the linear classifier designed for two-level multivariate data where the model is represented by the BCS covariance structure. The crucial advantage of this structure is its ability to capture two types of variations in the data, where the covariance matrix $\mathbf{\Sigma}_0$ represents variation of feature variables within observations, and $\mathbf{\Sigma}_1$ represents variation of feature variables between any two time replicates. Our main results are as follows:

• The asymptotic normality of the BCS classifier hinges on the distributional property of the matrix parameters $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$. We derive the pooled unbiased estimators of the matrix parameters of the BCS structure and show that their distributions are Wishart.

• Our high-dimensional asymptotic framework allows the number of replicates (u) to grow faster than the sample size, so that the total number of variables can be larger than the sample size.

• Main results of this paper can be extended to more general settings. In particular, two-level data can have many other covariance structures, e.g., separable covariance structure with half structured, half unstructured or both structured or both unstructured. An extension of the BCS classifier to more than two classes is also possible.

We extend the consideration of Kubokawa et al. (2013) to two-level data and derive the asymptotic approximation of EPMC in the setting which allows the number of replicates to grow with the weaker constrain $m < n$, unlike $mu < n$ in the case of Kubokawa et al. (2013). Our results are in line with Kubokawa et al. (2013) and reduces to their model exactly with $u = 1$.

To extend the measure of classification accuracy to the multi-class case, one can apply the *multiple binary comparisons* in p-populations classification problem as it is suggested in Wu (2003), or to use *one-against-all* approach applied in Dettling (2005) or to consider the distance based classifier, see Srivastava (2006).

Acknowledgement

Anuradha Roy gratefully acknowledges financial support in part from the Royal Swedish Academy of Sciences, Göran Gustafsson Stiftelse, grant 2012-2014, and Tatjana Pavlenko gratefully acknowledges financial support in part from the Swedish Research Council under grant 421-2008-1966. This work was completed while Anuradha Roy was visiting the KTH Royal Institute of Technology, Sweden in 2013.

References

[1] Anderson T. W. (2003). An introduction to multivariate statistical analysis. Third Ed. John Wiley & Sons, New Jersey.

[2] Bartlett M. S. (1951). An inverse matrix adjustment arising in discriminant analysis. Ann. Math. Statist 22(1): 107-111.

[3] Coelho, A. C. and Roy A. (2013). Testing of hypothesis of a block compound symmetric covariance matrix. The University Of Texas At San Antonio, College Of Business Working Paper Series, WP # 0012MSS-253 -2013.

[4] Dettling, M., Gabrielson, E. and Parmigiani G. (2005). Searching for Differentially Expressed Gene Combinations Genome Biology, 6, R88.

[5] Fujikoshi, Y., Ulyanov, V. V. and Shimizu R. (2010). Multivariate statistics : high-dimensional and large-sample approximations, Hobroken, New Jersey: Wiley.

[6] Fujikoshi, Y. and Seo, T. (1998). Asymptotic approximations for EPMC's of the linear and the quadratic discriminant functions when the sample sizes and the dimensions are large, Random Oper. Stochastic Equations 6, 269-280.

[7] Kubokawa, T. Hyodo, M. and Srivastava, M. S. (2013). Asymptotic expansion and estimation of EPCM for linear classification rules in high dimension. J. Multivariate Anal., 115, 496-515.

[8] Leiva, R. (2007). Linear discrimination with equicorrelated training vectors. Journal of Multivariate Analysis 98: 384-409.

[9] Leiva, R. and Roy, A. (2011). Linear Discrimination for Multi-level Multivariate Data with Separable Means and Jointly Equicorrelated Covariance Structure, Journal of Statistical Planning and Inference, 141(5), 1910-1924.

[10] Leiva R. and Roy A. (2012). Linear discrimination for three-level multivariate data with separable additive mean vector and doubly exchangeable covariance structure, Computational Statistics and Data Analysis, 56(6), 1644-1661.

[11] Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function. Ann. Math. Statist., 34, 1286-1301.

[12] Rao, C. R. (1945). Familial correlations or the multivariate generalizations of the intraclass correlation. Current Science 14: 66-67.

[13] Rao, C. R. (1953). Discriminant functions for genetic differentiation and selection. Sankhyā, 12: 229-246.

[14] Ritter G. and Gallegos M. T. (2002). Bayesian object identification: variants. Journal of Multivariate Analysis 81(2): 301-334.

[15] Roy, A. and Leiva R., Žežula I. and Daniel Klein (2013). Testing the Equality of Mean Vectors for Paired Doubly Multivariate Observations in Blocked Compound Symmetric Covariance Matrix Setup. Submitted.

[16] Roy, A. and Leiva R. (2013). Testing the Equality of Mean Vectors for Paired Doubly Multivariate Observations. University of Texas at San Antonio Working Paper Series # 0017MSS-253-2013.

[17] Roy, A. and Leiva R. (2011). Estimating and Testing a Structured Covariance Matrix for Three-level Multivariate Data, Communications in Statistics - Theory and Methods, 40(11), 1945-1963.

[18] Roy, A. and Leiva R. (2007). Discrimination with Jointly Equicorrelated Multi-level Multivariate Data, Advances in Data Analysis and Classification, 1(3), 175-199.

[19] Srivastava, M. S. (2006). Minimum distance classification rules for high dimensional data. J. Multivariate Anal. 97, 2057-2070.

[20] Srivastava, M. S. and Khatri, C. G. (1979). An introduction to multivariate statistics, North-Holland, New York.

[21] Wu, T. F., Lin, C. J. and Weng R. (2003) Probability estimates for multi-class classification by pairwise coupling. J. Mash. Learn. Res., 5, 975-1005.

[22] Wyman F. J., Young, D. M. and Turner, D. W. (1990) A comparison of asymptotic error rate expectations for the sample linear discriminant functions. Pattern Recognition, 23 775-783.

Table 1: Values of $\Phi(\gamma_0)$ for simulated data for different choices of $(n^{(1)} = n^{(2)})$, m, u, and $\mathcal{D}^2$.

$(n^{(1)}, n^{(2)}) \rightarrow$			(10,10)			(20,20)			(40,40)		
$\mathcal{D}^2 \downarrow$	$m \downarrow$	$u \rightarrow$	2	5	8	2	5	8	2	5	8
2	2		0.292	0.323	0.344	0.268	0.290	0.306	0.254	0.268	0.279
			(0.300)	*(0.334)*	*(0.354)*	*(0.270)*	*(0.294)*	*(0.312)*	*(0.255)*	*(0.269)*	*(0.281)*
	5		0.341	0.379 ‡	0.398 ‡	0.298	0.333	0.354 ‡	0.272	0.297	0.315
			(0.351)	*(0.387)*	*(0.405)*	*(0.302)*	*(0.338)*	*(0.359)*	*(0.273)*	*(0.300)*	*(0.318)*
	8		0.378	0.412 ‡	0.427 ‡	0.323	0.361 ‡	0.382 ‡	0.287	0.319	0.340
			(0.388)	*(0.419)*	*(0.433)*	*(0.327)*	*(0.366)*	*(0.386)*	*(0.289)*	*(0.322)*	*(0.343)*
	10		0.399 ‡	0.429 ‡	0.442 ‡	0.337	0.375 ‡	0.395 ‡	0.296	0.331	0.353 ‡
			(0.408)	*(0.435)*	*(0.447)*	*(0.342)*	*(0.380)*	*(0.399)*	*(0.298)*	*(0.334)*	*(0.355)*
4	2		0.201	0.227	0.247	0.180	0.195	0.208	0.169	0.178	0.185
			(0.210)	*(0.238)*	*(0.259)*	*(0.184)*	*(0.200)*	*(0.214)*	*(0.171)*	*(0.180)*	*(0.188)*
	5		0.252	0.292 ‡	0.318 ‡	0.206	0.236	0.258 ‡	0.183	0.201	0.216
			(0.262)	*(0.305)*	*(0.330)*	*(0.210)*	*(0.242)*	*(0.265)*	*(0.184)*	*(0.203)*	*(0.219)*
	8		0.299	0.342 ‡	0.365 ‡	0.230	0.269 ‡	0.294 ‡	0.195	0.221	0.242
			(0.310)	*(0.353)*	*(0.375)*	*(0.235)*	*(0.275)*	*(0.300)*	*(0.197)*	*(0.224)*	*(0.245)*
	10		0.330 ‡	0.369 ‡	0.390 ‡	0.246	0.287 ‡	0.314 ‡	0.204	0.234	0.256 ‡
			(0.340)	*(0.380)*	*(0.399)*	*(0.251)*	*(0.294)*	*(0.320)*	*(0.206)*	*(0.237)*	*(0.259)*

The notation ‡ represents failure of Kubokawa et al. (2013).

Table 2: Values of $\Phi(\gamma_0)$ for simulated data for different choices of $(n^{(1)} > n^{(2)})$, m, u, and $\mathcal{D}^2$

$(n^{(1)}, n^{(2)}) \rightarrow$			(12,8)			(25,15)			(50,30)		
$\mathcal{D}^2 \downarrow$	$m \downarrow$	$u \rightarrow$	2	5	8	2	5	8	2	5	8
2	2		0.278	0.292	0.298	0.258	0.267	0.272	0.249	0.256	0.260
			(0.288)	*(0.304)*	*(0.310)*	*(0.263)*	*(0.273)*	*(0.279)*	*(0.251)*	*(0.258)*	*(0.263)*
	5		0.312	0.322 ‡	0.320 ‡	0.276	0.286	0.287 ‡	0.260	0.270	0.275
			(0.324)	*(0.333)*	*(0.331)*	*(0.282)*	*(0.292)*	*(0.294)*	*(0.262)*	*(0.273)*	*(0.278)*
	8		0.342	0.343 ‡	0.337 ‡	0.291	0.296 ‡	0.293 ‡	0.269	0.279	0.282
			(0.353)	*(0.354)*	*(0.348)*	*(0.297)*	*(0.303)*	*(0.300)*	*(0.272)*	*(0.282)*	*(0.286)*
	10		0.361 ‡	0.359 ‡	0.350 ‡	0.300	0.302 ‡	0.296 ‡	0.274	0.284	0.285 ‡
			(0.372)	*(0.370)*	*(0.361)*	*(0.306)*	*(0.309)*	*(0.303)*	*(0.277)*	*(0.287)*	*(0.289)*
4	2		0.193	0.206	0.215	0.174	0.181	0.187	0.166	0.171	0.174
			(0.204)	*(0.219)*	*(0.229)*	*(0.180)*	*(0.188)*	*(0.194)*	*(0.169)*	*(0.173)*	*(0.177)*
	5		0.232	0.248 ‡	0.255 ‡	0.192	0.203	0.209 ‡	0.176	0.184	0.190
			(0.244)	*(0.262)*	*(0.268)*	*(0.198)*	*(0.211)*	*(0.217)*	*(0.178)*	*(0.187)*	*(0.193)*
	8		0.271	0.284 ‡	0.286 ‡	0.209	0.221 ‡	0.225 ‡	0.184	0.195	0.201
			(0.284)	*(0.298)*	*(0.299)*	*(0.216)*	*(0.228)*	*(0.232)*	*(0.187)*	*(0.198)*	*(0.205)*
	10		0.299 ‡	0.309 ‡	0.307 ‡	0.221	0.231 ‡	0.233 ‡	0.190	0.201	0.207 ‡
			(0.312)	*(0.322)*	*(0.321)*	*(0.227)*	*(0.239)*	*(0.241)*	*(0.193)*	*(0.205)*	*(0.211)*

The notation ‡ represents failure of Kubokawa et al. (2013).

Table 3: Values of $\Phi(\gamma_0)$ for simulated data for different choices of $(n^{(1)} < n^{(2)})$, m, u, and $\mathcal{D}^2$

$(n^{(1)}, n^{(2)}) \rightarrow$			(8,12)			(15,25)			(30,50)		
$\mathcal{D}^2 \downarrow$	$m \downarrow$	$u \rightarrow$	2	5	8	2	5	8	2	5	8
2	2		0.309	0.360	0.396	0.280	0.317	0.347	0.261	0.283	0.302
			(0.314)	*(0.368)*	*(0.403)*	*(0.280)*	*(0.321)*	*(0.352)*	*(0.260)*	*(0.283)*	*(0.303)*
	5		0.374	0.442 ‡	0.483 ‡	0.325	0.389	0.432 ‡	0.287	0.331	0.364
			(0.382)	*(0.448)*	*(0.486)*	*(0.327)*	*(0.393)*	*(0.435)*	*(0.287)*	*(0.332)*	*(0.367)*
	8		0.419	0.485 ‡	0.524 ‡	0.361	0.435 ‡	0.483 ‡	0.310	0.367	0.408
			(0.427)	*(0.489)*	*(0.524)*	*(0.364)*	*(0.439)*	*(0.485)*	*(0.310)*	*(0.369)*	*(0.410)*
	10		0.442 ‡	0.503 ‡	0.538 ‡	0.381	0.459 ‡	0.507 ‡	0.323	0.387	0.431 ‡
			(0.449)	*(0.505)*	*(0.537)*	*(0.385)*	*(0.462)*	*(0.508)*	*(0.325)*	*(0.390)*	*(0.433)*
4	2		0.212	0.253	0.286	0.187	0.213	0.236	0.173	0.187	0.200
			(0.219)	*(0.262)*	*(0.296)*	*(0.190)*	*(0.217)*	*(0.241)*	*(0.174)*	*(0.188)*	*(0.201)*
	5		0.276	0.345 ‡	0.393 ‡	0.223	0.277	0.320 ‡	0.192	0.223	0.251
			(0.284)	*(0.355)*	*(0.401)*	*(0.226)*	*(0.282)*	*(0.325)*	*(0.193)*	*(0.225)*	*(0.253)*
	8		0.332	0.407 ‡	0.455 ‡	0.257	0.328 ‡	0.381 ‡	0.210	0.255	0.293
			(0.340)	*(0.415)*	*(0.460)*	*(0.261)*	*(0.333)*	*(0.385)*	*(0.211)*	*(0.258)*	*(0.296)*
	10		0.365 ‡	0.437 ‡	0.482 ‡	0.278	0.357 ‡	0.412 ‡	0.221	0.275	0.318 ‡
			(0.373)	*(0.444)*	*(0.485)*	*(0.282)*	*(0.362)*	*(0.416)*	*(0.223)*	*(0.277)*	*(0.321)*

The notation ‡ represents failure of Kubokawa et al. (2013).

Table 4: Comparison of first and second order values of $\Phi(\gamma_0)$ in Kubokawa et al. (2013)

$\mathcal{D}^2$	m	$(n^{(1)}, n^{(2)})$	$\Phi(\gamma_0)$	
			Second order	First order
	10	(20, 20)	0.314	0.310
	10	(20, 20)	0.314	0.310
2	10	(30, 10)	0.269	0.265
	10	(10, 30)	0.379	0.377
	10	(20, 20)	0.226	0.221
4	10	(30, 10)	0.198	0.193
	10	(10, 30)	0.272	0.268
	50	(40, 40)	0.392	0.389
2	50	(50, 30)	0.359	0.356
	50	(30, 50)	0.429	0.427
	50	(20, 20)	0.324	0.319
4	50	(30, 10)	0.299	0.294
	50	(10, 30)	0.353	0.349